CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Dartmouth Digital Commons (Dartmouth College)

Dartmouth College
Dartmouth Digital Commons

Computer Science Technical Reports | Computer Science

3-20-2005

Department of Computer Science Activity 1998-2004

David Kotz
Dartmouth College

Follow this and additional works at: https://digitalcommons.dartmouth.edu/cs_tr

 Part of the Computer Sciences Commons

Dartmouth Digital Commons Citation

Kotz, David, "Department of Computer Science Activity 1998-2004" (2005). Computer Science Technical Report TR2005-534. https://digitalcommons.dartmouth.edu/cs_tr/267

This Technical Report is brought to you for free and open access by the Computer Science at Dartmouth Digital Commons. It has been accepted for inclusion in Computer Science Technical Reports by an authorized administrator of Dartmouth Digital Commons. For more information, please contact dartmouthdigitalcommons@groups.dartmouth.edu.

Department of Computer Science
Activity 1998–2004

David Kotz (editor)
dfk@dartmouth.edu

Technical Report TR2005-534
Department of Computer Science, Dartmouth College
http://www.cs.dartmouth.edu

March 20, 2005

Contents

1 Introduction

This report summarizes much of the research and teaching activity of the Department of Computer Science at Dartmouth College between late 1998 and late 2004. The material for this report was collected as part of the final report for NSF Institutional Infrastructure award EIA-9802068, which funded equipment and technical staff during that six-year period. This equipment and staff supported essentially all of the department's research activity during that period.

Indeed, the infrastructure helped to support and encourage collaboration among faculty in the department. In the individual project reports blow, note that many of the projects represent collaborations between two or more faculty. In some cases, the project involved collaboration with colleagues in other departments as diverse as Mathematics, Chemistry, Biology, Engineering, or Sociology.

Furthermore, the faculty in this department has a long tradition of integrating their research and educational activities. We involve countless undergraduate students in our research projects; whether as paid programmers or Senior Thesis students working for credit, these students learn a tremendous amount by working directly with faculty and graduate students. Furthermore, we have offered numerous special-topics courses related directly to the research underway, and created two new permanent courses on Robotics and on Computer Security.

The grant has also helped both existing and new faculty to succeed on a national and international scale. Many have won of prestigious awards and honors. NSF CAREER awards were granted to Hany Farid, Amit Chakrabarti, Sean Smith, and Chris Bailey-Kellogg. Sloan Foundation Fellowships have gone to Daniela Rus, Cliff Stein, Prasad Jayanti, Hany Farid, and Chris Bailey-Kellogg. Bruce Donald received a Guggenheim Fellowship and was appointed to a chaired professorship, the first for our department. Daniela Rus won a MacArthur Fellowship.

Our undergraduate and graduate students have also received honors. LeeAnn Tzeng, a graduate student, was awarded an NSF Graduate Fellowship. In 1999 the two CRA Outstanding Undergradate Awards went to Dartmouth students April Rasala and Marty Vona. Since that time seven of our undergraduates have received Honorable Mentions. Since 1999 sixty-nine of our undergraduates graduated with Honors for their undergraduate research.[1]

Finally, we are expanding. Last fall we moved into a new addition to our building, which increases our space by about 50%.

Please visit our department home page or the individual faculty web pages for more information.

[1] http://www.cs.dartmouth.edu/academics/honors.html.

2 Courses

Our faculty find many ways to mix teaching and research. Graduate and undergraduate students are involved directly in research, and research ideas tend to creep into core undergraduate and graduate courses, but perhaps the most measurable impact of our research on the curriculum is the series of "special topics" courses that are offered every year. Some are taught by regular faculty, and some are taught by visiting faculty; all enrich the experience for 5–20 students each time. Below is a list of the topics courses taught during the period of this award.

Linder, Spring 2004: *Rock-climbing: Design and Implementation of an Intelligent Robot*

Thompson, Spring 2004: *Information Retrieval*

McIlroy, Winter 2004: *Logic of Programming*

Donald, Winter 2004: *Computational Molecular Biology*

Cormen, Fall 2003: *How To Write, Evaluate, and Present Technical Papers in Computer Science*

Chakrabarti, Fall 2003: *Lower Bounds in Computer Science*

Jayanti, Winter 2003: *Theoretical Computer Science*

Donald, Fall 2002: *Algorithms in Computational Biology and Chemistry*

Hawblitzel, Fall 2002: *Types, Proofs, and Secure Systems*

Cormen, Winter 2003: *Parallel Computing*

Kotz, Winter 2003: *Pervasive Computing*

Aslam, Spring 2003: *Information Theory and Its Applications*

Makedon, Spring 2003: *Topics in Bio Medical Informatics: Algorithms, Tools, and Systems*

Drysdale, Spring 2002: *Computational Geometry*

Butler, Spring 2002: *Mobile Robots: Theory and Design*

Akay, Winter 2002: *Neural Networks and Computations*

Kotz, Winter 2002: *Context-Aware Mobile Computing*

Smith, Winter 2002: *Building and Breaking Secure Systems*

Liljenstam, Fall 2001: *Discrete Event Systems Simulation*

Jayanti, Fall 2001: *Theoretical Computer Science*

Smith, Spring 2001: *Building and Breaking Secure Systems*

Aslam, Winter 2001: *Machine Learning and Information Retrieval*

Stein, Winter 2001: *Network Flows*

Hawblitzel, Fall 2000: *Parallel and Distributed Simulation*

Donald, Spring 2000: *Topics in Computational Molecular Biology*

Farid, Spring 2000: *Fundamentals of Image and Signal Processing*

Jayanti, Winter 2000: *Distributed Algorithms*

Cormen, Winter 2000: *Parallel Computing*

Kotz, Fall 1999: *Context-Aware Mobile Computing*

Stein, Spring 1999: *Scheduling Algorithms*

Makedon, Spring 1999: *Computational Multimedia: Theory and Applications*

Donald, Winter 1999: *Topics in Computer Animation*

Kotz, Fall 1998: *Wireless Networks and Hand-held Computers*

3 Information Retrieval (Javed Aslam, Daniela Rus)

In the field of Information Retrieval, we have developed models, algorithms, and software for (1) using mobile agents for distributed information access, (2) automatically organizing vast quantities of digital data, (3) combining the results of multiple search engines to improve query retrieval, and (4) creating a unified framework for collectively solving the problems of metasearch, pooling, and system evaluation.

3.1 Activities and Findings

3.1.1 Mobile agents for information retrieval.

We used transportable agents primarily for distributed information access, in which a distributed collection of corpora is searched based on a query and the results extracted from each site are fused in a coherent picture. The main advantages of using agents in distributed information access are flexibility and performance. With agents, distributed collections can provide primitive operations rather than all possible search operations. An agent can combine these primitives into efficient, multi-step searches. By moving a small computation to the location of the data (with transportable agents), the network traffic and overall computation time is reduced.

We built information-access agents that interface with the well-known "Smart" information retrieval system. The Smart system is a successful statistical information-retrieval system that uses the vector-space model to measure the textual similarity between documents. The idea of the vector-space model is that each word that occurs in a collection defines an axis in the space of all words in the collection. A document is represented as a weighted vector in this space. The premise of this system is that documents that use the same words map to neighboring points and that statistics capture content similarity. Our "star" algorithm organizes a document collection into clusters that are naturally induced by the topic structure of collection, via a computationally efficient cover by dense subgraphs [APR04, APR00b, APR00a, ARR00, APR98b, APR98a, APR97]; the star algorithm is described in more detail in the following section.

Our data is a distributed collection of document repositories, each running an information-retrieval system. In our prototype, each collection consists of computer science technical reports. For a given query, an information agent visits a sequence of sites; at each site, it interacts with the local Smart agent to search the local collection. The results retrieved are brought home, or used as relevance feedback to refine the query. We consider the advantages and disadvantages of mobile agents for this sort of task, and develop planning algorithms suitable to minimize overhead [BGM+99].

We also conducted a series of related, but unpublished, experiments to measure the scalability and performance of persistent queries in a large document collection. Our Standing Query Server (SQS) received keyword-based queries from clients, performed an initial database search for the first 50 matching documents, clustered the resulting documents using an implementation of the "star clustering" algorithm mentioned above, and then stored the query. When new documents arrived in the database, the system ran each stored query over the new documents, and recalculated the document clusters. If a new document joined one of the clusters selected as most relevant by the client, and was above a certain relevance threshold, the client was notified and could retrieve the document. We ran experiments to discover the effect of the relevance threshold, database size, and number of standing queries on the performance of the system, and determined that the algorithm was consuming significant memory resources to cache associations. In fact, the memory used was more than the combined size of the documents themselves, leading to performance tradeoffs between the number of new documents added and the number of standing queries that could be supported. Overall, the performance did not appear to be practical for the quickly-growing document collections that we targeted.

3.1.2 Automatic information organization

The problem of automatically organizing vast quantities of electronic data is critical in our digital world. The amount of information stored in a typical library or on the World Wide Web, for instance, is immense and

growing at a phenomenal rate. In order to efficiently extract relevant pieces of information, users must be provided with sophisticated search mechanisms. Traditional query-based retrieval systems (such as Google and Alta Vista on the World Wide Web) allow the user to submit a query to which the system returns a collection of "relevant" documents. These systems often fail in the face of the vastness of the data in the system: often hundreds or thousands of documents are returned to the user, large numbers of which are irrelevant, forcing the user to manually search for the information required. Another approach is to provide the user with a pre-organized database (such as *Yahoo!* on the World Wide Web) through which the user can efficiently search. Unfortunately, these systems often require vast human resources to maintain, and they cannot hope to keep pace with the rate at which information is being accumulated.

To help alleviate this problem, we have developed algorithms and systems which can automatically organize vast quantities of digital data and assist users in searching through this data. Given an appropriate pairwise similarity function, our *star cover algorithm* can be used to automatically organize a collection of objects into tight clusters of mutually relevant objects [APR04, APR00b, APR00a, ARR00, APR98b, APR98a, APR97]. Unlike many previous algorithms which tended to be either inefficient or lacking in performance guarantees, the star cover algorithm is both *efficient* (linear time in the size of the input problem) and provably *accurate* (a lower bound on the pairwise similarity of intra-cluster objects is guaranteed in a geometric model of the information space). As a pre-processor, it can be used to automatically organize large databases of information; as a post-processor, it can be used to organize the results returned by a traditional query-based retrieval system. We further developed an *on-line* version of the star cover algorithm which can be used to automatically organize dynamic databases of information, such as the World Wide Web, and we developed a probabilistic model based on random graphs to show that the expected running time of this on-line algorithm is within lower order factors of the corresponding off-line algorithm [APR99, APR98b, APR00a]. Finally, we demonstrated that the star cover algorithm can be made scalable to large object collections through sampling techniques [ARR00]. The star cover algorithm has also proven useful in other domains such as document filtering [APR00b] and organizing data from functional magnetic resonance imaging studies.

3.1.3 Metasearch

Many search engines exist, both on the World Wide Web and for specific databases of information such as US patents and medical abstracts, to name but a few examples. No one search algorithm or technique works best over all types of queries and underlying databases. The problem of *metasearch* is to intelligently combine the results of multiple search engines in response to a given query, obtaining a overall response whose quality equals, or exceeds, that of the (unknown) best search engine. Many well-known and commonly used metasearch algorithms are based on *ad hoc* techniques and tend to be rather simplistic; for example, simply interleaving the ranked lists returned by the underlying search engines or ordering documents by their rank-sum over the underlying systems.

To combat this problem, we have developed rigorous, mathematical models for investigating the problem of metasearch. In one model, based on *decision theory*, the average performance of the underlying search engines can be assessed to obtain, for example, the expected probability that a document ranked at level i by search engine j will be relevant or irrelevant with respect to a given query. A collection of retrieved documents may then be ordered by *odds of relevance*, calculated in a Bayes-optimal manner by combining the probabilistic evidence of relevance obtained for each document from the underlying search engines [AM01, AM00].

We have implemented algorithms based on this decision-theoretic model and tested them against benchmark data from the TREC[2] competition. Even assuming search engine independence (a condition known not to hold in practice), the performance of his method equals or exceed the performance of the previously best, highly tuned, *ad hoc* techniques. Furthermore, unlike *ad hoc* techniques, this mathematical model provides a clear foundation for progress and further research.

[2]TREC, the Text REtrieval Conference, sponsored annually by the National Institute of Standards and Technology (NIST).

We have also studied and developed a model for metasearch based on *social choice theory*. The problem of metasearch can be viewed as a multi-candidate election: documents are "candidates" and search engines are "voters" expressing preferential rankings among the candidates. As such, algorithms from election theory can be brought to bear on the metasearch problem, leveraging on the large body of work developed in the field of social choice. Many interesting algorithmic issues arise in the analysis and implementation of metasearch algorithms based on social choice theory; for example, typical multi-candidate elections have a small number of candidates and a large number of voters while the analogous metasearch problem is quite the opposite. We have developed, analyzed and implemented a number of metasearch algorithms based on multi-candidate election strategies and tested them on benchmark data sets from the TREC competition and the World Wide Web [MA02, AM01]. Our results demonstrate that these algorithms are efficient and outperform widely used *ad hoc* techniques.

Finally, we have also demonstrated that metasearch improves the *consistency* of search systems across queries [MA01a], and we have developed a number of techniques for improving the quality of nearly all standard metasearch algorithms [MA01b], including the well-known and widely used *ad hoc* techniques.

3.1.4 Metasearch, Pooling, and System Evaluation

As a natural outgrowth of the research described above, we further (1) investigated the techniques of *metasearch*, *pooling* and *pseudo-evaluation*; (2) established formal connections between these seemingly disparate techniques; and (3) developed unified models for approaching all three tasks simultaneously. Individually, these techniques are used to either improve, or evaluate the effectiveness of, retrieval in the context of multiple available search strategies. By establishing formal connections between these techniques, we have developed formal frameworks for investigating all three techniques simultaneously with the end goals of (1) improving the quality of retrieval effectiveness in the context of multiple available search strategies, (2) improving the quality of the evaluation of search strategies, and (3) reducing the resources necessary to perform such evaluations.

As described above, *metasearch* is the process of fusing the ranked lists of documents returned by a collection of retrieval systems in response to a given query in order to obtain a combined list whose quality is better than any of the underlying lists. Many metasearch algorithms have been proposed and studied, and metasearch has proven to be an effective technique in improving retrieval quality.

Traditionally, retrieval systems are evaluated by constructing a test collection of documents, a set of query topics, and a set of relevance judgments for a subset of the collection (the *pool*) with respect to these topics. Much research has been conducted in how to best construct the pool of documents to be judged in order to effectively evaluate a collection of retrieval systems, and pooling techniques are often used to evaluate collections of retrieval systems in, for example, the annual TREC competitions.

Recently, Soboroff, Nicholas and Cahan surprisingly demonstrated that retrieval systems can be effectively ranked without relevance judgments by appropriately constructing a pool and assigning relevance judgments at random within the pool. This process of *pseudo-evaluation* holds out the promise of being able to efficiently extend the evaluation of retrieval systems to large collections of systems over large and possibly dynamic databases of information such as the World Wide Web.

While these techniques have generally been studied in isolation, we have shown there are deep connections between them; in effect, we have provided strong evidence that metasearch, pooling, and pseudo-evaluation are "three sides" of the same coin. To demonstrate this fact, we developed a general framework [APS03b, APS03a] which, given a user query and access to multiple retrieval systems, could (1) identify those documents returned by the underlying systems most likely to be relevant to the query (metasearch), (2) estimate the likely quality of each system's response to the query (pseudo-evaluation), and (3) identify those documents whose relevance, if known, would most likely influence the quality of the system evaluation (pooling). The end goal of this research is both theoretical and applied: first, to establish the formal connections among metasearch, pooling and pseudo-evaluation; and second, to develop highly effective *meta-retrieval* systems which could simultaneously perform metasearch, pseudo-evaluation and pooling, as well as incorporate user feedback in the form of relevance judgments for pooled documents to improve the quality of the above.

3.2 Products

Publications

[AF03] Javed A. Aslam and Meredith Frost. An information-theoretic measure for document similarity. In Jamie Callan, Gordon Cormack, Charles Clarke, David Hawking, and Alan Smeaton, editors, *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 449–450. ACM Press, July 2003.

[ALS01] Javed Aslam, Alain Leblanc, and Cliff Stein. Clustering data without prior knowledge. In Stefan Naher and Dorethea Wagner, editors, *Algorithm Engineering: 4th International Workshop*, volume 1982 of *Lecture Notes in Computer Science*, pages 74–86. Springer, 2001.

[AM00] Javed Aslam and Mark Montague. Bayes optimal metasearch: A probabilistic model for combining the results of multiple retrieval systems. In Nicholas J. Belkin, Peter Ingwersen, and Mun-Kew Leong, editors, *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 379–381. ACM Press, July 2000.

[AM01] Javed A. Aslam and Mark Montague. Models for metasearch. In W. Bruce Croft, David J. Harper, Donald H. Kraft, and Justin Zobel, editors, *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 276–284. ACM Press, September 2001.

[APR97] Javed Aslam, Katya Pelekhov, and Daniela Rus. Using high-quality clusters for summarizing and visualizing large document collections. In *The 1997 SIGIR Workshop on Summarization and Visualization for IR: Reducing the Information Overload*, July 1997.

[APR98a] Javed Aslam, Katya Pelekhov, and Daniela Rus. Generating, visualizing and evaluating high-quality clusters for information organization. In Ethan V. Munson, Charles Nicholas, and Derick Wood, editors, *Principles of Digital Document Processing: 4th International Workshop*, volume 1481 of *Lecture Notes in Computer Science*, pages 53–69. Springer, 1998.

[APR98b] Javed Aslam, Katya Pelekhov, and Daniela Rus. Static and dynamic information organization with star clusters. In George Gardarin, James French, Niki Pissinou, Kia Makki, and Luc Bouganim, editors, *Proceedings of the Seventh International Conference on Information and Knowledge Management*, pages 208–217. ACM Press, November 1998.

[APR99] Javed Aslam, Katya Pelekhov, and Daniela Rus. A practical clustering algorithm for static and dynamic information organization. In *Proceedings of the Tenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 51–60. ACM SIAM, January 1999.

[APR00a] Javed Aslam, Katya Pelekhov, and Daniela Rus. Information organization algorithms. In *Proceedings of the International Conference on Advances in Infrastructure for Electronic Business, Science, and Education on the Internet*, July 2000.

[APR00b] Javed Aslam, Katya Pelekhov, and Daniela Rus. Using star clusters for filtering. In Arvin Agah, Jamie Callan, and Elke Rundensteiner, editors, *Proceedings of the Ninth International Conference on Information Knowledge Management*, pages 306–313. ACM Press, November 2000.

[APR04] Javed A. Aslam, Ekaterina Pelekhov, and Daniela Rus. The star clustering algorithm for static and dynamic information organization. *Journal of Graph Algorithms and Applications*, 8(1):95–129, 2004.

[APS03a] Javed A. Aslam, Virgiliu Pavlu, and Robert Savell. A unified model for metasearch and the efficient evaluation of retrieval systems via the hedge algorithm. In Jamie Callan, Gordon Cormack, Charles Clarke, David Hawking, and Alan Smeaton, editors, *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 393–394. ACM Press, July 2003.

[APS03b] Javed A. Aslam, Virgiliu Pavlu, and Robert Savell. A unified model for metasearch, pooling, and system evaluation. In Ophir Frieder, Joachim Hammer, Sajda Quershi, and Len Seligman, editors, *Proceedings of the Twelfth International Conference on Information and Knowledge Management*, pages 484–491. ACM Press, November 2003.

[ARR00] Jay Aslam, Fred Reiss, and Daniela Rus. Scalable information organization. In *Proceedings of the Conference on Content-Based Multimedia Information Access*, pages 1033–1042, Paris, France, April 2000. C.I.D.-C.A.S.I.S.

[AS03] Javed A. Aslam and Robert Savell. On the effectiveness of evaluating retrieval systems in the absence of relevance judgments. In Jamie Callan, Gordon Cormack, Charles Clarke, David Hawking, and Alan Smeaton, editors, *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 361–362. ACM Press, July 2003.

[BGM+99] Brian Brewington, Robert Gray, Katsuhiro Moizumi, David Kotz, George Cybenko, and Daniela Rus. Mobile agents for distributed information retrieval. In Matthias Klusch, editor, *Intelligent Information Agents*, chapter 15, pages 355–395. Springer-Verlag, 1999.

[MA01a] Mark Montague and Javed A. Aslam. Metasearch consistency. In W. Bruce Croft, David J. Harper, Donald H. Kraft, and Justin Zobel, editors, *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 386–387. ACM Press, September 2001.

[MA01b] Mark Montague and Javed A. Aslam. Relevance score normalization for metasearch. In Henrique Paques, Ling Liu, and David Grossman, editors, *Proceedings of the Tenth International Conference on Information and Knowledge Management*, pages 427–433. ACM Press, November 2001.

[MA02] Mark Montague and Javed A. Aslam. Condorcet fusion for improved retrieval. In Konstantinos Kalpakis, Nazli Goharian, and David Grossman, editors, *Proceedings of the Eleventh International Conference on Information and Knowledge Management*, pages 538–548. ACM Press, November 2002.

3.3 Contributions

We contributed new models, algorithms, and software for solving a number of important problems in the field of Information Retrieval. We developed the infrastructure of mobile agents and demonstrated that they could be used to solve the problem of distributed information retrieval. We developed a new model and a number of efficient algorithms for automatically organizing data, and we have shown that this algorithm is both highly accurate and efficient for its stated purpose and that it can be used to help solve a number of other longstanding problems in the field of Information Retrieval, such as information filtering and persistent query retrieval. We developed the first theoretically grounded models and algorithms for metasearch, and we demonstrated that these new algorithms equal or exceed the performance of well known, highly tuned, and largely *ad hoc* methods. Finally, we developed a unified framework for solving the longstanding problems of metasearch, pooling, and system evaluation, demonstrating for the first time that these three seemingly disparate problems are in fact "three sides" of the same coin.

Human resources. We have trained five graduate students and seven undergraduate students as part of these projects. Two of the graduate students have completed their PhDs, and four of the undergraduate students have gone on to graduate school.

4 Algorithms and Theory (Amit Chakrabarti)

Amit Chakrabarti worked on computational complexity theory and, in a separate project, on algorithms for NP-hard optimization problems.

4.1 Activities and Findings

4.1.1 Research

Chakrabarti's research in complexity theory focused on the theory of *lower bounds*, which are provable limitations on our ability to minimize resources in solving computational problems. Typical resources of interest include computational time, space and inter-computer communication, the latter being a primary focus of Chakrabarti's work. A short while before joining Dartmouth he and others introduced the notion of *information complexity* which measures the cost of communication using information theoretic ideas originally developed by Shannon. The advantage is that this new measure is amenable to mathematical analysis and can often be related to the traditional measure of number of bits communicated.

Chakrabarti and others used the information complexity technique to show optimal communication lower bounds for the problem of multiparty SET-DISJOINTNESS [2]. This lower bound has a significant consequence in the field of memory-efficient data stream computations: it can be shown to imply optimal space lower bounds on single pass and multipass algorithms for computing frequency moments.

Chakrabarti, along with a coauthor, also gave the first nontrivial lower bound for high dimensional approximate nearest neighbour searching that works even in the case of randomized algorithms [3]. He also developed an algorithm to match the lower bound on the hypercube, thereby proving the optimality of both bounds. These results close a long line of research work spanning a decade and involving over two dozen researchers worldwide. Chakrabarti is now focusing on extending the results to more challenging metric spaces and computational models.

A secondary research activity has been developing approximation algorithms for NP-hard optimization problems. In this area, Chakrabarti has helped develop the first constant-factor approximation algorithms for the unsplittlable flow problem (a basic problem in network optimization) on line and ring networks [1]. The techniques developed there were subsequently used by other researchers to extend the result to trees. This work also addressed the unsplittable flow problem on expander graphs.

4.1.2 Teaching

At Dartmouth, Chakrabarti has developed and taught a new course on "Lower Bounds in Computer Science." The courses is designed to give beginning graduate students (and advanced undergraduates) a gentle introduction to lower bound techniques for a variety of problems studied in complexity theory. A website for the course contains a detailed day-by-day plan, homework excercises and solutions [4]. The second installment of this course will lead to lecture notes that will be made available publicly on the Web to benefit other teachers.

4.1.3 Outreach Activities

Since Dartmouth's Computer Science Department is small in size, contact with researchers at other institutions is essential for creating a sense of belonging to the larger CS community. To help with this, Chakrabarti started the Dartmouth Theory Seminar series [5], which invites speakers with exciting new theoretical results to Dartmouth to speak about their work, so that Dartmouth's students and faculty with an interest in Theory could keep in touch with developments related to but outside of their own research focus. This series is now in its third term and is expected to continue on a regular basis.

4.2 Products

Publications

[1] Amit Chakrabarti, Chandra Chekuri, Anupam Gupta, and Amit Kumar. Approximation algorithms for the unsplittable flow problem. *Algorithmica*, 2005. To appear.

[2] Amit Chakrabarti, Subhash Khot, and Xiaodong Sun. Near-optimal lower bounds on the multi-party communication complexity of set disjointness. In *Proc. 18th Annual IEEE Conference on Computational Complexity*, pages 107–117, 2003.

[3] Amit Chakrabarti and Oded Regev. An optimal randomised cell probe lower bound for approximate nearest neighbour searching. In *Proc. 45th Annual IEEE Symposium on Foundations of Computer Science*, 2004. to appear.

[4] Amit Chakrabarti. Lower Bounds in Computer Science: course plan, homeworks and solutions. http://www.cs.dartmouth.edu/~ac/Teach/CS85-Fall03/, December 2003.

[5] Amit Chakrabarti. Dartmouth Theory Seminars: current and previous terms. http://www.cs.dartmouth.edu/~ac/DTS/, regularly updated.

4.3 Contributions

The concept of information complexity, though introduced before Chakrabarti joined Dartmouth, is being brought to fruition as part of his work here. This concept has already proved to be very useful in a diverse range of problems in complexity theory and is expected to continue to yield new exciting results. Chakrabarti's own optimal results about high dimensional approximate nearest neighbour searching are a good example.

The work on approximation algorithms for unsplittable flow has influenced other researchers in their work extending the results, as discussed above.

The Dartmouth Theory Seminars brings together graduate students in large numbers and exposes them to new theoretical ideas in a way that no course at Dartmouth currently does.

The larger goal of any theoretical work remains the same as in eras past: to deepen humanity's understanding of the fundamental process of computing.

5 Out-of-Core Computing (Thomas Cormen)

Professor Thomas H. Cormen and his students have developed algorithms that work with massive data, along with software infrastructure to promote algorithm engineering of such algorithms.

5.1 Activities and Findings

When the amount of data exceeds the capacity of main memory (archaically known as "core" memory), the data must reside outside of main memory. We have focused on computing platforms in which the data reside on one or more disks connected to one or more nodes of a distributed-memory cluster. Rather than the disks hanging off a common I/O bus, we assume that each disk is connected to only a given node.

Out-of-core algorithms

We have designed and implemented several variations on out-of-core algorithms, focusing on two problems in particular: sorting and Fast Fourier Transform (FFT).

FFT. Prior to the start of this award, we had developed out-of-core algorithms for one-dimensional FFTs on a uniprocessor with multiple disks and for one-dimensional FFTs on a distributed-memory cluster. During the award period, we extended our results to multidimensional FFTs on a distributed-memory cluster. We examined both the dimensional method [BC99], which works on one dimension at a time, and the vector-radix method [Rin01], which works on a small piece of all dimensions at once. We designed and implemented efficient out-of-core algorithms for each of these cases. Subsequently, we answered theoretical questions about how to order and group the individual dimensions in the dimensional method [Fin01].

Sorting. Prior to the start of this award, we had begun to investigate methods for out-of-core sorting on distributed-memory clusters. During the award period, we made great strides in designing and implementing oblivious algorithms for out-of-core sorting. An oblivious sorting algorithm is one in which the sequence of operations is independent of the values being sorted. The advantage of an oblivious algorithm in out-of-core sorting is that the entire sequence of disk I/O and interprocessor communication operations is known in advance; hence, an implementation can overlap computation, communication, and I/O in order to mitigate the costs of these operations.

We developed several out-of-core sorting algorithms based on the oblivious columnsort algorithm of Leighton [Lei85]. Our first implementation [CCW01] used asynchronous I/O calls and static scheduling to overlap I/O with computation and communication. It sorted in four passes over the data, subject to an upper limit on the number of records being sorted. (Each pass reads each record once and writes each record once.)

Our next implementations [CC02] used the standard pthreads interface to fully redesign the structure of each pass. We constructed each pass as multiple stages of a software pipeline, passing buffers from stage to stage. Each stage was a thread. This structure offered two significant advantages over the first implemention. One was that the structure of the code simplified because each stage operated synchronously rather than asynchronously. The second advantage was that computation, communication, and I/O operations were scheduled dynamically, rather than statically, by the thread scheduler. The resulting code was both simpler and significantly faster than the first implementation. In the same paper, we showed how to reduce the number of passes from four down to three, with no change in the upper limit on problem size.

Subsequent work showed how to increase the upper limit on the number of records to sort. We did so by two different approaches. One was an algorithm-engineering view, in which we took a different interpretation of what a column in columnsort comprises. The second approach was an algorithmic approach, in which we devised several modifications to columnsort to relax the problem-size bound on the algorithm itself, with the relaxed bound carrying over directly to the out-of-core adaptation. This work appears in [Cha04, CC03, CC04, CCH, CHC03].

Software infrastructure

In our work on out-of-core columnsort, we observed that each pass was actually a software pipeline, in which stages mapped to threads and buffers traverse the pipeline. We found that programming the details of thread creation and destruction, as well as the coordination among threads (via semaphores), was tedious and difficult to debug, however.

To simplify the task of developing asynchronous software pipelines that pass buffers among stages, we developed middleware that we call Asynchronous Buffered Computation Design and Engineering Framework Generator (ABCDEFG), or FG for short. In an FG pipeline, while one stage accesses high-latency data, such as data on disk, other stages can make progress.

Each stage of an FG pipeline is a synchronous function written in C++. A stage accepts a buffer from its predecessor stage, operates on that buffer, and conveys the buffer to its successor stage. Buffers come from a pool of buffers allocated by FG at the start of pipeline execution. FG-added source and sink stages recycle buffers back through the pipeline so that the buffer pool may be much smaller than the total amount of data being processed and so that the program can avoid the overhead inherent in repeated calls to `new` and `dispose`. (These C++ operations are particularly expensive in multithreaded programs, such as those produced by FG.)

The programmer specifies the pipeline structure by using basic features of C++. An object represents each stage. An object also represents each thread, but if each stage maps to its own thread, the programmer need not represent any threads; FG does so on its own. The pipeline structure appears in a null-terminated array of stage objects. The programmer also specifies which C++ function each stage runs, how many buffers to create, and the size of each buffer. FG handles the rest. It creates and destroys threads, runs the pipeline, and handles all synchronization between adjacent stages.

We have found that programs written in FG are smaller, simpler, and faster than their counterparts written with static asynchronous scheduling or even with explicit pthreads calls [CD04, DC05b].

Newer releases of FG provide even more capability [DC05a]. We include optimizations to improve performance. Stages may be replicated, either statically by the programmer or dynamically by FG itself. FG also now alters thread priorities to use resources more efficiently; again, this action may be initiated by either the programmer or FG. FG now also suits programs that do not fit a simple, linear pipeline structure. To extend the range of computations that fit into its framework, FG now incorporates fork-join and DAG structures. Not only do these structures allow for more programs to be designed for FG, but they also can enable significant performance improvements over linear pipeline structures.

Other activities

In addition to the main activities described above, we have engaged in the following:

- We finished the second edition of the textbook *Introduction to Algorithms* [CLRS01].
- We produced an instructor's manual [CLL02] for the second edition of the textbook.
- We released a LaTeX macro package [Cor03a] for pseudocode in *Introduction to Algorithms*, and we published Java implementations [Cor03b] of the algorithms in the book.
- We implemented Rajasekaran's out-of-core sorting algorithm [Raj01] on a shared-memory system [Pea99].

Education activities

The work outlined above aided in the education of several undergraduate and graduate students at Dartmouth College.

- Geeta Chaudhry received her Ph.D. in 2004. Her thesis [Cha04] was on out-of-core sorting, and she contributed substantially to the FG project. Chaudhry is currently a postdoctoral fellow at Dartmouth, and she is continuing her line of research on out-of-core sorting.

- Elena Riccio Davidson is in her third year of Ph.D. studies. She has performed the great majority of the design, implementation, and evaluation of FG. This project will form the basis of her Ph.D. thesis, which we expect to be completed by June 2007.

- Wei Zhang is in his second year of Ph.D. studies. He is attempting to use FG to reduce latency in cache-aware algorithms.

- Michael Ringenburg received his M.S. degree in 2001. His thesis [Rin01] was on the vector-radix method and out-of-core FFTs. Ringenburg is currently a Ph.D. student in the Computer Science program at the University of Washington.

- Lauren Baptist's senior honors thesis [Bap99] was on the dimensional method for multidimensional, multiprocessor, out-of-core FFTs. She received departmental honors for this work. Baptist subsequently received an S.M. degree from MIT in 2000, and she has been a software developer at Google since 2000.

- Jeremy Fineman's senior honors thesis [Fin01] answered several theoretical questions about the dimensional method in an out-of-core setting. He received departmental honors for this work. Fineman is currently a Ph.D. student in Computer Science at MIT.

- Elizabeth Hamon's senior honors thesis [Ham03] produced the first working implementation of FG. Hamon received departmental high honors for this work. She has been a software developer at Google since 2003.

- Matthew Pearson's senior honors thesis [Pea99] was an implementation of Rajasekaran's shared-memory, out-of-core sorting algorithm. Pearson received department honors for this work. He became a software developer at Sun Microsystems in 1999, but we do not know if he is still with Sun.

- Clara Lee and Erica Lin, both Dartmouth undergraduate Computer Science majors in the Class of 2003, coauthored the instructor's manual for the second edition of *Introduction to Algorithms* [CLL02]. This project required them to learn the conventions of textbook writing and production. They did a job so remarkable that they received full credit as coauthors. Lee is currently a software developer at Google. Lin's current employment is unknown.

- Jessica Webster's senior honors thesis, under the aegis of Dartmouth's Mathematics and Social Sciences Program, was the development of a laptop-runnable program to record every event in a baseball game and to collect and produce statistics over any period of time. She received program honors for this work. Webster subsequently received a master's degree in human-computer interaction at Tufts University. She is currently employed in private industry.

- Tiffany Wong's senior honors thesis [Won01a, Won01b] had two parts. In one part, she developed a tool for a debugging project. In the other part, she provided database support for Webster's baseball-scoring program. Wong received departmental honors for her work. She received her J.D. in 2004 from the University of Chicago Law School and is currently an associate at Wilson Sonsini Goodrich & Rosati, where she practices corporate and securities law.

- A. Cristina Maracine's senior honors thesis was on the interaction of caching and FG. She received departmental honors for her work.

- Brunn Roysden's senior honors thesis was on thread priorities and FG. His work led to some of the newer features of FG. Roysden received departmental honors for his work.

In addition, the textbook *Introduction to Algorithms*, Second Edition, is widely used. Just shy of 90,000 copies are in print as of June 2004.

5.2 Products

The project has produced the following.

Research publications:

- Refereed journal publication: [CCH].
- Refereed conference publications: [BC99, CC02, CCW01, CD04, DC05a, DC05b].
- Ph.D. thesis: [Cha04].
- Master's thesis: [Rin01].
- Senior honors theses: [Bap99, Fin01, Ham03, Pea99, Won01a, Won01b].
- Article in submission: [CC04].
- Technical report: [CC03].
- Tutorial and reference manual: [DCb].
- Other: [CHC03, CDC03].

Books: [CLL02, CLRS01].

Software: [Cor03a, Cor03b].

Web site and software: [DCb].

Publications

[Bap99] Lauren M. Baptist. Two algorithms for performing multidimensional, multiprocesso r, out-of-core FFTs. Technical Report PCS-TR99-350, Dartmouth College Department of Computer Science, June 1999.

[BC99] Lauren M. Baptist and Thomas H. Cormen. Multidimensional, multiprocessor, out-of-core FFTs with distributed memory and parallel disks. In *Proceedings of the Eleventh Annual ACM Symposium on Parallel Algorithms and Architectures*, pages 242–250, June 1999.

[CC02] Geeta Chaudhry and Thomas H. Cormen. Getting more from out-of-core columnsort. In *4th Workshop on Algorithm Engineering and Experiments (ALENEX 02)*, pages 143–154, January 2002.

[CC03] Geeta Chaudhry and Thomas H. Cormen. Stupid columnsort tricks. Technical Report TR2003-444, Dartmouth College Department of Computer Science, April 2003.

[CC04] Geeta Chaudhry and Thomas H. Cormen. Slabpose columnsort: A new oblivious algorithm for out-of-core sorting on distributed-memory clusters. Submitted to *Algorithmica*, 2004.

[CCH] Geeta Chaudhry, Thomas H. Cormen, and Elizabeth A. Hamon. Parallel out-of-core sorting: The third way. *Cluster Computing*. To appear.

[CCW01] Geeta Chaudhry, Thomas H. Cormen, and Leonard F. Wisniewski. Columnsort lives! An efficient out-of-core sorting program. In *Proceedings of the Thirteenth Annual ACM Symposium on Parallel Algorithms and Architectures*, pages 169–178, July 2001.

[CD04] Thomas H. Cormen and Elena Riccio Davidson. FG: A framework generator for hiding latency in parallel programs running on clusters. In *Proceedings of the 17th International Conference on Parallel and Distributed Computing Systems (PDCS-2004)*, pages 137–144, September 2004.

[CDC03] Thomas H. Cormen, Elena Riccio Davidson, and Siddhartha Chatterjee. Asynchronous buffered computation design and engineering framework generator (ABCDEFG). In *High-End Computing Revitalization Task Force Workshop*, June 2003. White paper.

[Cha04] Geeta Chaudhry. *Parallel Out-of-Core Sorting: The Third Way*. PhD thesis, Dartmouth College, 2004.

[CHC03] Geeta Chaudhry, Elizabeth A. Hamon, and Thomas H. Cormen. Relaxing the problem-size bound for out-of-core columnsort. In *Proceedings of the Fifteenth Annual ACM Symposium on Parallel Algorithms and Architectures*, June 2003. SPAA Revue paper.

[CLL02] Thomas H. Cormen, Clara Lee, and Erica Lin. *Instructor's Manual to Accompany Introduction to Algorithms, Second Edition*. The MIT Press and McGraw-Hill, 2002. Available only by contacting the publishers.

[CLRS01] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms*. The MIT Press and McGraw-Hill, second edition, 2001.

[Cor03a] Thomas H. Cormen. The clrscode package for LaTeX2e. http://www.cs.dartmouth.edu/~thc/clrscode/, June 2003.

[Cor03b] Thomas H. Cormen. Java implementations of algorithms from *Introduction to Algorithms*, second edition. http://www.cs.dartmouth.edu/~thc/clrsjava/ (password-protected website), October 2003.

[DCa] Elena Riccio Davidson and Thomas H. Cormen. *Asynchronous Buffered Computation Design and Engineering Framework Generator (ABCDEFG): Tutorial and Reference*. Dartmouth College Department of Computer Science. Available at http://www.cs.dartmouth.edu/FG/.

[DCb] Elena Riccio Davidson and Thomas H. Cormen. FG. http://www.cs.dartmouth.edu/FG/.

[DC05a] Elena Riccio Davidson and Thomas H. Cormen. Building on a framework: Using FG for more flexibility and improved performance in parallel programs. In *19th IEEE International Parallel and Distributed Processing Symposium (IPDPS 2005)*, April 2005. To appear.

[DC05b] Elena Riccio Davidson and Thomas H. Cormen. The FG programming environment: Reducing source code size for parallel programs running on clusters. In *Second Workshop on Productivity and Performance in High-End Computing (P-PHEC)*, February 2005.

[Fin01] Jeremy Fineman. Optimizing the dimensional method for performing multidimensional, multiprocessor, out-of-core FFTs. Technical Report TR2001-402, Dartmouth College Department of Computer Science, June 2001.

[Ham03] Elizabeth A. Hamon. Enhancing asynchronous parallel computing. Technical Report TR2003-460, Dartmouth College Department of Computer Science, 2003.

[Lei85] Tom Leighton. Tight bounds on the complexity of parallel sorting. *IEEE Transactions on Computers*, C-34(4):344–354, April 1985.

[Pea99] Matthew D. Pearson. Fast out-of-core sorting on parallel disk systems. Technical Report PCS-TR99-351, Dartmouth College Department of Computer Science, 1999.

[Raj01] Sanguthevar Rajasekaran. A framework for simple sorting algorithms on parallel disk systems. *Theory of Computing Systems*, 34(2):101–114, 2001.

[Rin01] Michael F. Ringenburg. Applying the vector radix method to multidimensional, multiprocessor, out-of-core fast Fourier transforms. Master's thesis, Dartmouth College Department of Computer Science, March 2001. Also Dartmouth College Department of Computer Science Technical Report TR2001-388.

[Won01a] Tiffany M. Wong. An implementation of object-oriented program transformation for thought-guided debugging. Technical Report TR2001-395, Dartmouth College Department of Computer Science, 2001.

[Won01b] Tiffany M. Wong. Implementing a database information system for an electronic baseball scorecard. Technical Report TR2001-396, Dartmouth College Department of Computer Science, 2001.

5.3 Contributions

The project has contributed the following to the principal discipline of the project:

- A new understanding of the value of oblivious algorithms in high-latency computing, specifically, out-of-core computing.
- Designs and implementations of fast, robust sorting algorithms that run on distributed-memory clusters.
- Designs and implementations of out-of-core programs for one-dimensional and multidimensional FFTs on distributed-memory clusters.
- A software infrastructure consisting of middleware that promotes the development of software that makes efficient use of resources in a high-latency environment.

The project has contributed the following to the development of human resources:

- Geeta Chaudhry received her Ph.D. from work on the project.
- Michael Ringenburg received his M.S. degree from work on the project.
- Elena Riccio Davidson and Wei Zhang are in the process of Ph.D. research work on the project.
- Lauren Baptist, Jeremy Fineman, Elizabeth Hamon, A. Cristina Maracine, Matthew Pearson, Brunn Roysden, Jessica Webster, and Tiffany Wong all completed senior honors thesis on the project.
- Clara Lee and Erica Lin coauthored an instructor's manual on the project.
- Finally, Thomas H. Cormen collaborated with and supervised all of the above, and he was promoted to full Professor, while working on the project.

6 Computational Biology (Bruce Donald)

Bruce Donald's group did extensive work on planning and object recognition algorithms for structural molecular biology.

6.1 Activities and Findings

A wealth of interesting computational problems arise in developing and applying information technology to understand the molecular machinery of the cell. We began to develop high-throughput, automated systems for structural and functional genomics. In particular, our goal is to design algorithms that require data measurements of only a few key biophysical parameters, which will be obtained from fast, minimal, cheap experiments. We utilize two classes of biophysical experiment: nuclear magnetic resonance (NMR) and mass spectrometry (MS). MS is highly amenable to automation and promises to scale to macromolecular complexes, but yields limited structural information. NMR yields a great deal of structure content, but is not easily automated. We began to bridge this gap through sophisticated information technology, obtaining more structural content from MS and bringing more automation to NMR. Our approach will allow the structure and function of biopolymers to be assayed at a fraction of the time and cost of current methods.

Structural genomics — computing three-dimensional protein structure — is a vital component of modern biology. Current techniques for computing structure using NMR spectroscopy require dozens of experiments and months of spectrometer time. We began to extend and adapt algorithms from computer vision, signal processing, and robotics in a system for determining structure from only a few key NMR spectra. We have begun by developing novel techniques for computing protein secondary structure and main-chain assignment from just four spectra. In our preliminary work, we represent NMR data with graphs encoding potential interactions between amino acid residues, and apply object recognition-like graph algorithms to uncover subgraphs encoding secondary structure, overcoming a 5-10% signal-to-noise ratio. We then use a probabilistic algorithmic framework to align identified secondary structure to the primary sequence. We plan to extend this system to a fully automated system for computing biopolymer structure from protein and DNA NMR data, using techniques from Expectation/Maximization, polyspectral analysis, and geometric reasoning to analyze input spectra, and graph algorithms to search for and exploit structural constraints. The novelty in our approach devolves to a minimalist emphasis designed for high-throughput assays, the employment of new techniques from computer science, and the application of computational methods from robotics, planning, and machine vision. JIGSAW views structure determination in noisy data as a form of *object recognition* and leverages algorithms from robotics and machine vision. It demonstrates two key insights: *graph-based secondary structure pattern discovery*, and *assignment by alignment*.

Structure is an integral part of function: an important task in functional genomics is the discovery of structure-activity relations (SAR), indicating binding modes of protein-protein and DNA-protein complexes. We began to develop techniques for determining structure-activity relations using new computational protocols for MS. SAR by MS enzymatically cleaves a cross-linked complex and analyzes the resulting mass spectrum for mass peaks of hypothesized fragments. Depending on the binding mode, some cleavage sites will be shielded; the absence of anticipated peaks for the corresponding fragments implicates those fragments as either part of the interaction region or inaccessible due to conformational change upon binding. Thus different mass spectra provide evidence for different structure-activity relations. An important problem in SAR by MS is the potential for mass degeneracy: when two possible fragments have approximately the same mass, the existence of one or the other cannot be uniquely inferred from a mass peak. Thus we began to study automated planning techniques, in the style of active sensing, that seek to maximize the potential information content of MS experiments and efficiently interpret the resulting data. These planning techniques determine isotopic mass manipulations that will yield unambiguous mass peaks for the resulting fragments. In preliminary work, we have proved the optimization form of the problem to be NP-complete, have tested a randomized planning algorithm and probabilistic framework for estimating interpretability, and have developed an efficient multi-spectral data analysis technique. We plan to apply computational models to predict binding modes, incorporate additional MS analysis

techniques, extend the experiment planner to plan protease/endonuclease selection, and further validate and refine our techniques. We view experiment planning as a form of active sensing (from vision and robotics), where the the "controls" are selective isotopic labeling and limited proteolysis, and "sensing" is done by MS. This problem can be cast as a combinatorial optimization problem, with the objective of finding a labeling strategy that minimizes the amount of mass degeneracy (spectral overlap).

Fast structure determination and SAR by MS deliver complementary information, which we want to synergistically integrate into a complete high-throughput, automated system for structural and functional genomics. For example, a preliminary structure obtained from NMR data constrains the hypotheses on potential fragment interactions, allowing SAR by MS to focus on a smaller set of masses to disambiguate. Alternatively, SAR by MS can guide structure computation with spatial proximity information derived from binding mode. The information content in limited proteolysis provides additional synergy: potential cleavage sites in regular secondary structure have a smaller likelihood of being cleaved. Hence, secondary structure determination can guide prediction of fragments and interpretation of mass peaks, or vice-versa.

Our work integrates research and training of young scientists in the interdisciplinary field of computational molecular biology. We believe that information technology can provide leverage in structural and functional genomics, comparable to the impact of computational methods in gene sequencing and mapping. High-throughput automated structure and function will increase our understanding of biopolymer interactions in systems of significant biochemical and pharmacological interest, on a genomic scale. Our work will yield valuable results both in computational science, with the development and analysis of new algorithms, and in biochemistry, with computational systems useful for high-throughput structural and functional assays. This is a very fruitful area for the application of the robotics and vision techniques developed in our laboratory.

Two papers on the NMR work appear at the prestigious *International Conference on Computational Molecular Biology (RECOMB'2000 and 2001)*. Another, on the MS work, appeared at the *International Conference on Intelligent Systems for Molecular Biology* in August, 2000. We also published two journal papers in the *Journal of Computational Biology*. One is on our NMR work, the other is on our Mass Spec work.

Based on this work, I was awarded a John Simon Guggenheim Memorial fellowship for 2001-2001 for my proposal entitled "New Frontiers in Physical Geometric Algorithms," to apply robotics and vision algorithms to the challenges of structural proteomics.

6.1.1 Education

Donald developed a new Course, CS 88/188, "Topics in Computational Molecular Biology."

> *'Strictly speaking, molecular biology is not a new discipline, but rather a new way of looking at organisms as reservoirs and transmitters of information. This new vision opened up possibilities of action and intervention that were revealed during the growth of genetic engineering.'*
>
> - Michel Morange, "A History of Molecular Biology," Harvard University Press (1998).
>
> Our goal is to look at some algorithmic problems related to three-dimensional structures in chemistry and molecular biology, emphasizing the perspective of geometric algorithms. We hope to consider a variety of topics (guided by the interests of the participants), and to make the seminar interesting to people with as wide a range of backgrounds as possible.
>
> Some of the topics we may cover include: Protein and RNA-folding, Distance Geometry and Assignment for Protein NMR, DNA arrays, the Phase Problem in X-ray Crystallography, Rational Drug Design, Molecular Docking, identifying structural domains and motifs in proteins, and conformational search. We may read several papers on structural genomics, and papers on mass spectrometry for functional genomics.
>
> The CS-Bio seminar is open to graduate students, and advanced undergraduates with a background in both algorithms and systems (at least CS 25 and CS 23). A background in biology is useful

but not required. Students should be interested in doing some outside reading in biochemistry and biophysics. Students will be required to present papers in the seminar, and to do a project. Non-CS students (e.g., in biology and chemistry) with an interest in computational issues are invited as well.

Course is entirely on the Web.
See http://www.cs.dartmouth.edu/~brd/Teaching/.

6.2 Products

We published a number of papers, including three in *Algorithmica* (arguably the most prestigious algorithms journal in computer science), and one in SIGGRAPH (the most selective and best refereed conference in computer science). Finally, my book, "Algorithmic and Computational Robotics: New Directions", was published by A. K. Peters. Details are provided below.

Papers in Refereed Journals

1. "A Novel Ensemble-Based Scoring and Search Algorithm for Protein Redesign, and its Application to Modify the Substrate Specificity of the Gramicidin Synthetase A Phenylalanine Adenylation Enzyme," (with R. Lilien, B. Stevens, and A. Anderson), *Journal of Computational Biology*, (In press) 2004.

2. "A Probability-Based Similarity Measure for Saupe Alignment Tensors with Applications to Residual Dipolar Couplings in NMR Structural Biology," (with A. Yan and C. Langmead) The *International Journal of Robotics Research* Special Issue on Robotics Techniques Applied to Computational Biology, (In press) 2004.

3. "Exact Solutions for Internuclear Vectors and Backbone Dihedral Angles from NH Residual Dipolar Couplings in Two Media, and Their Application in a Systematic Search Algorithm for Determining Protein Backbone Structure" (with L. Wang), *Journal of Biomolecular NMR* 2004; 29(**3**):223–242.

4. "An Expectation/Maximization Nuclear Vector Replacement Algorithm for Automated NMR Resonance Assignments" (with C. Langmead), *Journal of Biomolecular NMR* 2004; 29(**2**):111–138.

5. "A Polynomial-Time Nuclear Vector Replacement Algorithm for bAutomated NMR Resonance Assignments," (with C. Langmead, A. Yan, R. Lilien, and L. Wang) *Journal of Computational Biology* 2004; 11(**2–3**):277–298.

6. "A Subgroup Algorithm to Identify Cross-Rotation Peaks Consistent with Non-Crystallographic Symmetry" (with R. Lilien, C. Bailey-Kellogg, and A. Anderson). *Acta Crystallographica D: Biological Crystallography* 2004; D**60**, 1057–1067.

7. "The Crystal Structure of Dihydrofolate Reductase-Thymidylate Synthase from *Cryptosporidium hominis* Reveals a Novel Architecture for the Bifunctional Enzyme," (with R. O'Neil, R. Lilien, R. Stroud, and A. Anderson) *Journal of Eukaryotic Microbiology* 2003; 50(**6**):555-556. Cover article.

8. "Phylogenetic Classification of Protozoa Based on the Structure of the Linker Domain in the Bifunctional Enzyme, Dihydrofolate Reductase-Thymidylate Synthase," (with R. O'Neil, R. Lilien, R. Stroud, and A. Anderson) *Journal of Biological Chemistry* 2003; 278(**52**):52980–52987.

9. "Probabilistic Disease Classification of Expression-Dependent Proteomic Data from Mass Spectrometry of Human Serum," (with R. Lilien and H. Farid), *Journal of Computational Biology*, 10(**6**) 2003 pp. 925-946.

10. "Phase-Independent Rhythmic Analysis of Genome-Wide Expression Patterns", (with C. Langmead, A. Yan, and C. R. McClung), *Journal of Computational Biology* 10(**3-4**) 2003, pp. 521-536.

11. "Reducing Mass Degeneracy in SAR by MS (Structure-Activity Relation by Mass Spectrometry) by Stable Isotopic Labeling" (with C. Bailey-Kellogg, J. Kelley, and C. Stein) *Journal of Computational Biology* 8(**1**):19–36, 2001.

12. "The NOESY Jigsaw: Automated Protein Secondary Structure and Main-Chain Assignment from Sparse, Unassigned NMR Data," (with C. Bailey-Kellogg, A. Widge, J. Kelley, M. Berardi, and J. Bushweller), *Journal of Computational Biology,* 7**(3-4)** (2000) pp. 537–558.

Papers in Refereed Conferences

1. "Algorithmic Challenges in Structural Molecular Biology and Proteomics," Sixth International Workshop on the Algorithmic Foundations of Robotics (WAFR), Utrecht/Zeist, The Netherlands. July 11-13, 2004 pp. 1–10 (Invited).

2. "High-Throughput 3D Structural Homology Detection via NMR Resonance Assignment" (with C. Langmead), the *IEEE Computational Systems Bioinformatics Conference (CSB)*, Stanford CA, pp. 278–289 (August, 2004).

3. "Analysis of a Systematic Search-Based Algorithm for Determining Protein Backbone Structure from a Minimal Number of Residual Dipolar Couplings" (with L. Wang), the *IEEE Computational Systems Bioinformatics Conference (CSB)*, Stanford CA, pp. 319–330 (August, 2004).

4. "A Novel Ensemble-Based Scoring and Search Algorithm for Protein Redesign, and its Application to Modify the Substrate Specificity of the Gramicidin Synthetase A Phenylalanine Adenylation Enzyme," (with R. Lilien, B. Stevens, and A. Anderson), *Proceedings of the Eighth Annual International Conference on Research in Computational Molecular Biology (RECOMB),* San Diego (March 27-31, 2004) pp. 46-57.

5. "3D Structural Homology Detection via Unassigned Residual Dipolar Couplings," (with C. Langmead), *Proc. IEEE Computational Systems Bioinformatics Conference (CSB)*, Stanford CA. (August 10, 2003) pp. 209-217. ISBN 0-7695-2000-6.

6. "A Polynomial-Time Nuclear Vector Replacement Algorithm for Automated NMR Resonance Assignments," (with C. Langmead, A. Yan, R. Lilien, and L. Wang), *Proceedings of the Seventh Annual International Conference on Research in Computational Molecular Biology (RECOMB),* Berlin (2003) pp. 176–187.

7. "A Maximum Entropy Algorithm for Rhythmic Analysis of Genome-Wide Expression Patterns", (with C. Langmead and C. R. McClung), Proc. IEEE Computer Society Bioinformatics Conference (CSB), Stanford University, Palo Alto (August 14–16, 2002) pp. 237–245.

8. "Phase-Independent Rhythmic Analysis of Genome-Wide Expression Patterns", (with C. Langmead, T. Yan, and C. R. McClung), Proceedings of The Sixth Annual International Conference on Research in Computational Molecular Biology (RECOMB), Washington DC (2002). pp. 205–215.

9. "Extracting Structural Information Using Time-Frequency Analysis of Protein NMR Data" (with C. Langmead), The Fifth Annual International Conference on Research in Computational Molecular Biology (RECOMB), Montreal, April 22–25 (2001) pp. 164–175.

10. "Physical Geometric Algorithms for Structural Molecular Biology" (with C. Bailey-Kellogg, J. Kelley, and R. Lilien) invited paper at the Special Session on Computational Biology & Chemistry, Proc. IEEE International Conference on Robotics and Automation (ICRA) 2001, Seoul, Korea, pp. 940-947.

11. "Reducing Mass Degeneracy in SAR by MS (Structure-Activity Relation by Mass Spectrometry) by Stable Isotopic Labeling" (with C. Bailey-Kellogg, J. Kelley, and C. Stein) *Proc.* 8^{th} *International Conference on Intelligent Systems for Molecular Biology (ISMB'2000)*, AAAI Press, (August 20–23, 2000) La Jolla, CA, pp. 13–24.

12. "The NOESY Jigsaw: Automated Protein Secondary Structure and Main-Chain Assignment from Sparse, Unassigned NMR Data," (with C. Bailey-Kellogg, A. Widge, J. Kelley, M. Berardi, and J. Bushweller), The Fourth Annual International Conference on Research in Computational Molecular Biology (RECOMB), Tokyo, Japan, April 8 — 11, 2000, pp. 33–44.

Conference Abstracts and Posters

1. "High-Throughput 3D Homology Detection via NMR Resonance Assignment" (with C. Langmead), *Eighth Annual International Conference on Research in Computational Molecular Biology (RECOMB),* San Diego (March 27-31, 2004). In *Currents in Computational Molecular Biology, 2004* (ed. A. Gramada and P. Bourne) p. 522.

2. A Probability-Based Similarity Measure for Saupe Alignment Tensors with Applications to Residual Dipolar Couplings in NMR Structural Biology (with A. Yan and C. Langmead) *Eighth Annual International Conference on Research in Computational Molecular Biology (RECOMB),* San Diego (March 27-31, 2004). In *Currents in Computational Molecular Biology, 2004* (ed. A. Gramada and P. Bourne) pp. 437-438.

3. A Framework for Automated NMR Resonance Assignments and 3D Structural Homology Detection (with C. Langmead). Gordon Conference on Computational Methods in Biomolecular NMR, Ventura, CA, January 18-23, 2004.

4. Exact Solutions for Internuclear Vectors and Dihedral Angles from Two RDCs and Their Application in a Systematic Search Algorithm for Determining Protein Backbone Structure (with L. Wang, R. Mettu, R. Lilien, and A. Yan). Gordon Conference on Computational Methods in Biomolecular NMR, Ventura, CA, January 18-23, 2004.

5. "An Exact Algorithm For Determining Protein Backbone Structure From NH Residual Dipolar Couplings," (with L. Wang, R. Mettu, and R. Lilien), *Proc. IEEE Computer Society Bioinformatics Conference (CSB)*, Stanford University, Palo Alto, (August 10, 2003) pp. 611-612. ISBN 0-7695-2000-6. (Refereed).

   Winner of Best Poster Award.

6. "Ensembles of active site conformations allow structure-based redesign and drug design," (with A. Anderson, R. Lilien and V. Popov), 225^{th} American Chemical Society National Meeting, March 23-28, 2003. New Orleans, LA.

7. "Modeling Protein Flexibility for Structure-Based Active Site Redesign," (with R. Lilien and A. Anderson) The Sixth Annual International Conference on Research in Computational Molecular Biology (RECOMB), Washington DC (2002). In *Currents in Computational Molecular Biology, 2002* (ed. L. Florea et al.) pp 122-123. (Refereed).

8. "Modeling of Protein Flexibility for Computational Active Site Redesign," (with R. Lilien and A. Anderson) the 16th Annual National MD/PhD Conference, Given Institute , Aspen, Colorado, July 13-15, 2001.

9. "Computational Screening Studies for Core-Binding Factor Beta (CBF-β): Use of Multiple Conformations to Model Receptor Flexibility," (with R. Lilien, M. Sridharan, X. Huang, and J. Bushweller), the 8^{th} International Conference on Intelligent Systems for Molecular Biology (ISMB'2000), (August 20–23, 2000) La Jolla, CA.

10. "Time-Frequency Analysis of Protein NMR Data" (with C. J. Langmead), the 8^{th} International Conference on Intelligent Systems for Molecular Biology (ISMB'2000), (August 20–23, 2000) La Jolla, CA.

11. "The NOESY Jigsaw: Automated Protein Secondary Structure and Main-Chain Assignment from Sparse, Unassigned NMR Data," (with C. Bailey-Kellogg, A. Widge, J. J. Kelley, III, M. J. Berardi, and J. H. Bushweller), the 8^{th} International Conference on Intelligent Systems for Molecular Biology (ISMB'2000), (August 20–23, 2000) La Jolla, CA.

Lectures and Colloquia

1. Workshop on Geometry in NMR Protein Structure Determination and NMR Structural Biology, Bellairs Research Institute of McGill University, Holetown, Barbados. Three talks:

   "Computational Challenges in NMR Structural Genomics," January 14, 2005.

   "An Expectation/Maximization Nuclear Vector Replacement Algorithm for Automated NMR Resonance Assignments," January 14, 2005.

   "Automated Protein Structure and Assignment from Sparse, Unassigned NMR Data," January 15, 2005.

2. "Protein Geometry and its Role in Structural Molecular Biology and Proteomics," Meeting of the American Mathematical Society (AMS), Lawrenceville, New Jersey, April 17-18, 2004.

3. "Algorithmic Challenges in Structural Molecular Biology and Proteomics." Given at:

   Biomedical Engineering Department, University of Michigan, Ann Arbor. January 27, 2005.

   Computer Science Department, University of Chicago. November 10, 2004.

   Plenary lecture, Sixth International Workshop on the Algorithmic Foundations of Robotics (WAFR), Utrecht/Zeist, The Netherlands. July 11, 2004,

   Dartmouth Computer Science Department, July 7, 2004,

   Computer Science Department, Tufts University, April 13, 2004, and

   MIT Computer Science and Artificial Intelligence Laboratory (CSAIL), Feb. 19, 2004.

4. "An Expectation/Maximization Nuclear Vector Replacement Algorithm for Automated NMR Resonance Assignments," Harvard Medical School, November 20, 2003.

5. "Computational Biochemistry," Foley Inaugural Lecture, Dartmouth College, October 8, 2003.

6. "Algorithmic Challenges in Structural Genomics," Penn Bioinformatics Forum, University of Pennsylvania, April 24, 2003.

7. "Computational Challenges in NMR Structural Genomics," Dartmouth Physics Department Colloquium, May 24, 2002.

8. "Algorithmic Challenges in Structural Molecular Biology," Robert Mueller-Thuns Distinguished Lecture in Computer Science, at the University of Illinois (Urbana-Champaign), March 11-12, 2002.

9. "Algorithmic Challenges in Structural Molecular Biology," Triangle Distinguished Lecture in Computer Science, University of North Carolina at Chapel Hill, Duke, and N.C. State, Feb. 18-19, 2002.
10. "Use of Multiple Conformations to Model Protein Flexibility in Core-Binding Factor," Department of Biochemistry, Dartmouth Medical School, March 4, 2002.
11. "Algorithms and Systems for High-Throughput NMR Structural Molecular Biology," Harvard University, Department of Engineering and Applies Scienced, December 14, 2001.
12. "Algorithms and Systems for High-Throughput Structural Molecular Biology," Sandia National Labs, March 15, 2001.
13. "Algorithms and Systems for High-Throughput Structural Molecular Biology," at *The New Biology: Technologies for Resolving Macromolecular Communications,* An International Symposium Sponsored by the Association of Biomolecular Resource Facilities, Feb. 25, 2001 (San Diego).
14. "Algorithms for High-Throughput Structural Molecular Biology," The Rowland Institute for Science, Cambridge, MA, October 25, 2000.
15. "Algorithms for Structural Molecular Biology" Harvard Medical School, September 19, 2000.
16. "Motion Planning Opportunities in Haptics and Structural Biology," EU-NSF Workshop on Motion Planning, Laboratoire d'Analyse et d'Architecture des Systèmes, Centre National de la Recherche Scientifique (LAAS-CNRS), Toulouse, France, June, 2000.

6.3 Contributions

The contributions are discussed above.

Human resources. I have worked with several students and postdoctoral fellows during this period.

- Chris Langmead, Ph.D. Now a professor at CMU.
- Ryan Lilien, Ph.D/M.D. student.
- Chris Bailey-Kellogg, postdoc. After four years as professor at Purdue, rejoined Dartmouth as Assistant Professor.
- Jack Kelly, postdoc. Now a professor at Michigan.
- Tim Danford, undergraduate.
- Alik Widge, undergraduate. Now at CMU.
- 4 other undergraduate students.

7 MEMS (Bruce Donald)

Bruce Donald has an active research program involving Microelectromechanical Systems (MEMS).

7.1 Activities and Findings

7.1.1 Algorithms for Sensorless Manipulation Using a Vibrating Surface

Existing industrial parts feeders move parts through a sequence of mechanical filters that reject parts in unwanted orientations. These feeders require the design of specialized devices such as baffles, cutouts, nests, or traps for each part. In a paper published in *Algorithmica* (one of three), we described a programmable apparatus that uses a vibrating surface for sensorless, non-prehensile manipulation, where parts are systematically positioned and oriented without sensor feedback or force closure. The idea is to generate and change the dynamic modes of a vibrating surface. Depending on the node shapes of the surface, the position and orientation of the parts can be actively controlled. Our research goal is to develop a science base for manipulation using programmable force fields.

The vibrating surface creates a two-dimensional force vector field. By chaining together sequences of force fields, the equilibrium states of a part in the field can be cascaded to obtain a desired final state. We describe efficient polynomial-time algorithms that generate sequences of force fields for sensorless positioning and orienting of planar parts, and we show that these strategies are complete. Finally we consider parts feeders that can only implement a finite set of force fields. We show how to plan and execute strategies for these devices, and discuss the tradeoff between mechanical complexity and planning complexity.

7.1.2 A Single Universal Force Field Can Uniquely Pose Any Part up to Symmetry

Recent work in parts handling advocates the investigation of a new generation of devices for parts feeding, sorting, positioning, and assembly. Unlike robot grippers, conveyor belts, or vibratory bowl feeders, these devices generate force fields in which the parts move until they may reach a stable equilibrium pose.

The development of the theory of programmable force fields has yielded a number of strategies to uniquely position and orient parts. Typically, more than one fields are applied in sequence to achieve the desired result. In our *IEEE Transactions on Robotics and Automation* paper in 2000, we show that unique part poses can be achieved with a single field. In particular, we present a single field that positions and orients any non-symmetric part into two stable equilibrium poses. Then we show that for any laminar part there exists a field in which the non-symmetric part reaches a unique stable equilibrium pose. Our latter result leads to the design of devices that can act as "universal parts feeders" proving an earlier conjecture about their existence.

Recent research in the theory of programmable force fields has yielded open-loop strategies to uniquely position, orient, and sort parts. These strategies typically consist of several fields that have to be employed in sequence to achieve a desired final pose. The length of the sequence depends on the complexity of the part.

In our paper, we show that unique part poses can be achieved with just one field. First, we exhibit a single field that positions and orients any laminar part (with the exception of certain symmetric parts) into two stable equilibrium poses. Then we show that for any laminar part there exists a field in which the part reaches a unique stable equilibrium pose (again with the exception of symmetric parts). Our second result leads to the design of "universal parts feeders", proving an earlier conjecture about their existence. We argue that universal parts feeders are relatively easy to build.

7.1.3 Untethered Scratch Drive Actuators

Joint work: Bruce Donald, Chris Levey, Craig McGray, Daniela Rus

Objectives. The goal of this research is to enable autonomous locomotion at the micro-scale. To achieve this with existing microfabrication technologies, we are investigating the development of scratch drive actuators that can operate in an untethered fashion. We are fabricating these devices using the MUMPS (Multi-User MEMS Process) fabrication process from Cronos Integrated Microsystems (formerly MCNC).

Our untethered scratch drives comprise two novel systems that are currently under development. The first is a capacitive power couple for delivering power to devices that are not physically wired to the substrate. This will allow the devices to locomote in an untethered fashion. The second novel system in our untethered scratch drives is the self-release mechanism by which our devices cut their ties with the fabrication substrate.

In order to fabricate these devices, we need a layer of insulation between two layers of conductive silicon. The MUMPS process does not provide an insulating layer that will work in this regard. So, we are developing ways of post-processing MUMPS devices so as to introduce an intermediate layer of insulation after the fact.

Technical Review - Scratch Drive Actuators A scratch drive is a direct-drive actuator that operates through electrostatic attraction. It is composed of a thin silicon plate with a bushing at the front end. The plate is typically in the range of 80 microns long and wide, and 1.4 microns thick. The bushing height is typically in the 1-2 micron range.

The scratch drive operates as follows. When a voltage is applied between the silicon plate and the substrate beneath it, the plate is drawn down into contact with the substrate. Since the front of the plate is supported by the bushing, strain energy is stored in the plate, and the edge of the bushing is pushed forwards. When the voltage is removed, the strain is released and the scratch drive plate moves forwards.

When an AC signal is applied, the above cycle is continuously repeated, and the scratch drive moves forward in a step-wise manner. Scratch drives have been operated at frequencies as low as 10 Hz, and as high as 2 kHz, with voltages ranging from 30 - 200 V.

Technical Approach - Capacitive Couple. The scratch drives that we are building do not require a direct electrical connection (wire) in order to apply the drive voltage. These devices receive power from an underlying grid of electrodes that do not restrict the movement of the drive. This enables us to study the locomotion of the drive when it is not guided by physical tethers or rails.

The electrodes are arranged such that given any two adjacent electrodes, one has positive voltage and the other has negative voltage. Each electrode is smaller than a scratch drive, so that no matter the position or orientation of a given scratch drive, it always lies above some area of positive voltage, and some area of negative voltage.

Because the scratch drive is made of conductive silicon but is coated with an insulating layer, charge flows within the scratch drive in response to the voltage on the underlying electrodes, but it will not flow into the electrodes. So, there is charge build-up on the underside of the scratch drive plate. This charge build-up causes the electrostatic attraction that results in motion of the scratch drive.

The typical drive voltage of a scratch drive actuator is 150 V . In order to achieve an equivalent voltage with the capacitive couple, approximately 300 volts must be applied between the high-voltage electrodes and the grounded electrodes.

Technical Approach - Release Mechanism The scratch drive devices must remain attached to their underlying substrate throughout the fabrication process. However, they must be detached from the substrate prior to operation. So, a release mechanism is required. We would like this release mechanism to be compatible with batch fabrication. i.e. the devices should be able to self-release.

To accomplish this, the scratch drives are originally suspended above the substrate by a long thin beam. The beam is intentionally flawed where it joins the scratch drive plate. When voltage is first applied to the electrodes beneath the scratch drive, the electrostatic attraction draws the device towards the substrate. The resulting deformation stress in the beam is concentrated at the intentional flaw. The beam then snaps at the flaw and the device is released.

Technical Approach - MUMPS Post-Processing MUMPS is a three-layer polysilicon fabrication process with a top layer of metal. (The metal layer is incompatible with our post-processing steps, and is therefore not used in our devices.) The silicon layers are shaped and separated by intervening sacrificial layers of silicon dioxide. The oxide layers are removed with a hydrofluoric acid wet etch.

In our devices, the sacrificial oxide separates the scratch drives from the underlying electrodes. When the oxide is removed, the drives are suspended about 2 microns above the electrodes. We need to add a layer of insulator that is thick enough to prevent dielectric breakdown at the driving voltage of approximately 200 V, but that is not so thick that it fuses the scratch drives to the electrodes.

To do this, we grow thermal oxide on the silicon by heating in water vapor at 700° C. (The relatively low oxidation temperature prevents thermal stress from deforming the thin scratch drive plates.) Since this process coats all silicon surfaces in insulator, there is then no way to make electrical contact with the substrate. So, the devices are lithographically patterned with a very low-resolution mask, and holes are opened up in the oxide above electrical contact pads.

Status. This research is currently in-process. The scratch drives have been designed and processed through MUMPS. Mechanical stability of the devices has been verified. We are currently characterizing the process of applying the intermediate insulating layer. We look forward to verifying the power couple, the release mechanism, and the forward-drive process.

7.2 Products

Publications

1. "A Single Universal Force Field Can Uniquely Orient Non-symmetric Parts," (with K.-F. Böhringer, F. Lamiraux, and L. Kavraki), in *Robotics Research*, eds. J. Hollerbach and D. Koditschek, Springer-Verlag (London) 2000, pp. 395-402.

2. "CMOS Integrated Organic Ciliary Actuator Arrays for General-Purpose Micromanipulation Tasks," (with J. Suh, R. B. Darling, K.-F. Böhringer, H. Baltes, and G. Kovacs), in *Distributed Manipulation*, ed. K. Böhringer et al., Kluwer Academic Publishing (2000), pp. 191–216.

3. "A Distributed, Universal Device for Planar Parts Feeding: Unique Part Orientation in Programmable Force Fields," (with K.-F. Böhringer, F. Lamiraux, and L. Kavraki), in *Distributed Manipulation*, ed. K. Böhringer et al., Kluwer Academic Publishing (2000), pp. 1–28.

4. "Power Delivery and Locomotion of Untethered Micro-Actuators," (with C. Levey, C. McGray, D. Rus, and M. Sinclair) *Journal of Microelectromechanical Systems*, (2003) 10**(6)**:947–959.

5. "Algorithmic MEMS," (with K.-F. Böhringer) in *Robotics: The Algorithmic Perspective*, ed. P. Agarwal, L. Kavraki, and M. Mason, A. K. Peters, Natick, MA (1998). pp. 1–20.

6. Using Programmable Vector Fields," (with K.-F. Böhringer, F. Lamiraux, and L. Kavraki), IEEE Transactions on Robotics and Automation, 16(2), April 2000, pp. 157-170.

7. Surface," (with K.-F. Böhringer, V. Bhatt, and K. Goldberg) *Algorithmica*, 26**(3/4)**, March/April (2000). Special Issue on Algorithmic Foundations of Robotics, pp. 389–429.

8. Micromanipulation Tool for Small Objects," (with J. Suh, R. B. Darling, K.-F. Böhringer, H. Baltes, and G. Kovacs), *Journal of Microelectromechanical Systems,* vol. 8, No. 4 (Dec. 1999), pp. 483–496.

9. K.-F. Böhringer and D. Halperin) *Discrete and Computational Geometry*, vol. 22 (1999), pp. 269–285.

10. with Applications to MEMS Actuator Arrays and Vibratory Parts Feeders," (with K.-F. Böhringer and N. C. MacDonald), *International Journal of Robotics Research,* vol. 18(2), February, 1999 pp. 168–200.

11. "Untethered Micro-Actuators for Autonomous Micro-robot Locomotion: Design, Fabrication, Control, and Performance" (with C. Levey, C. McGray, D. Rus, and M. Sinclair). 11th *International Symposium of Robotics Research (ISRR)*, October 19-22, 2003, Siena, Italy.

12. "Power Delivery and Locomotion of Untethered Micro-Actuators," (with C. Levey, C. McGray, D. Rus, and M. Sinclair) *IEEE MEMS (Proc. IEEE International Conference on Micro Electro Mechanical Systems)*, Kyoto, Japan (January 19-23, 2003) pp. 124-129.

13. "Fully Programmable MEMS Ciliary Actuator Arrays for Micromanipulation Tasks," (with J. Suh, R. B. Darling, K.-F. Böhringer, H. Baltes, and G. Kovacs), Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), San Francisco (April, 2000) pp. 1101–1108.

14. "A Single Universal Force Field can Uniquely Pose any Part up to Symmetry," (with K.-F. Böhringer, F. Lamiraux, and L. Kavraki), 9th *International Symposium of Robotics Research (ISRR)*, Snowbird, Utah, October 9-12 (1999).

15. *Algorithmic MEMS*, (with K.-F. Böhringer), 3rd *International Workshop on the Algorithmic Foundations of Robotics (WAFR),* Houston, TX (1998).

Invited papers

1. "CMOS Integrated Organic Ciliary Array as a General-Purpose Micromanipulation Tool," (with J. Suh, R. B. Darling, K.-F. Böhringer, H. Baltes, and G. Kovacs), IEEE International Conference on Robotics and Automation, Workshop on Distributed Manipulation (Detroit) May 11, 1999.

2. "Part Orientation with One or Two Stable Equilibria Using Programmable Vector Fields," (with K.-F. Böhringer, F. Lamiraux, and L. Kavraki), IEEE International Conference on Robotics and Automation, Workshop on Distributed Manipulation (Detroit) May 11, 1999.

3. "Micro Contacts and Micro Manipulation with MEMS Actuator Arrays" (with K.-F. Böhringer) IEEE International Conference on Robotics and Automation, Workshop on Modeling, Contact Analysis, and Simulation of Mechanical Systems in Robotics and Manufacturing, (Belgium, 1998).

Lectures and Colloquia

1. "MEMS Algorithms and Systems for Massively-Parallel Distributed Manipulation," GRASP Lab, Department of Computer Science, University of Pennsylvania, April 25, 2003.

2. "MEMS Algorithms and Systems for Massively-Parallel Distributed Manipulation," Middlebury College, November 14, 2000.

3. "MEMS Algorithms and Systems for Massively-Parallel Distributed Manipulation," Department of Mechanical Engineering, University of Michigan, September 29, 2000.

4. "MEMS Algorithms and Systems for Massively-Parallel Distributed Manipulation," NSF Planning Meeting, Ninth International Symposium on Robotics Research, Snowbird, Utah, October 10, 1999.

7.3 Contributions

The contributions are discussed above.

8 Robotics (Bruce Donald)

Professor Donald has a long history of work in robotics.

8.1 Activities and Findings

8.1.1 Mobile Robot Self-Localization Without Explicit Landmarks

Localization is the process of determining the robot's location within its environment. More precisely, it is a procedure which takes as input a geometric map, a current estimate of the robot's pose, and sensor readings, and produces as output an improved estimate of the robot's current pose (position and orientation). We describe a combinatorially precise algorithm which performs mobile robot localization using a geometric model of the world and a point-and-shoot ranging device. We also describe a rasterized version of this algorithm which we've implemented on a real mobile robot equipped with a laser rangefinder we designed. Both versions of the algorithm allow for uncertainty in the data returned by the range sensor. We also present experimental results for the rasterized algorithm, obtained using our mobile robots. This work is described in one of three papers we published in *Algorithmica*.

8.1.2 Using Haptic Vector Fields for Animation Motion Control

We are developing paradigms and algorithms for browsing and editing families of animations using a haptic force-feedback device called a Phantom. These techniques may be generalized to navigation of any high degree-of-freedom system from a lower degree-of-freedom control space, with applications to telerobotics and simulation of virtual humans. We believe that modeling the animation configuration space coupled with the highly interactive nature of the haptic device provides us with useful and intuitive means of control.

We have implemented our ideas in a system for the manipulation of animation motion capture data; in particular, anthropomorphic figures with 57 degrees of freedom are controlled by the user in real time. We treat trajectories, which encode animation, as first-class objects; haptic manipulation of these trajectories results in change to the animation. We have several haptic editing modes in which these trajectories are either haptically deformed or performed by the user with expressive control subject to dynamic haptic constraints. The initial trajectories are given by sample animations (for example, motion capture data) but may be authored by other means. This work is described in the IEEE International Conference on Robotics and Automation (ICRA) 2000.

8.1.3 Constrained Prehensile Manipulation: Distributed Manipulation with Ropes

Our 2000 ICRA paper paper describe new experiments with a distributed manipulation system. We study a system in which multiple robots cooperate to move large objects such as furniture and boxes using a constrained prehensile manipulation mode, by wrapping ropes around them. The system consists of three manipulation skills: tieing ropes around objects, affecting translations using a flossing manipulation gait, and affecting rotations using a ratcheting manipulation gait. We present experimental data and discuss the non-holonomic nature of this system.

Manipulating objects with ropes has several advantages over the direct manipulation of an object using a team of robots. First, the method allows for the parallel and synchronized manipulation of multiple objects. (That is, several objects can be wrapped together). Second, wrapping a rope around an object permits *constrained prehensile manipulation* with non-prehensile robots. The rope can conform to any object geometry on-line, in that the same wrapping protocol works for all geometries and there is no need to specify a model in advance. Finally, the rope can be viewed as a tool that allows the robot system to exert torques that are larger than the torque limit of each individual robot. Our notion of constrained prehensile manipulation is different than the classical definition of prehensile manipulation, which usually denotes a force closure grasp. In *constrained prehensile manipulation,* perturbations along any differential direction can be resisted (in our case, due to the taut

rope) but the object can be moved only along certain lower dimensional degrees of freedom. Thus, constrained prehensile manipulation systems are *non-holonomic.*

8.1.4 Visibility-Based Planning of Sensor Control Strategies

In another paper in *Algorithmica*, we consider the problem of planning sensor control strategies that enable a sensor to be automatically configured for robot tasks. In this paper we present robust and efficient algorithms for computing the regions from which a sensor has unobstructed or partially obstructed views of a target in a goal. We apply these algorithms to the Error Detection and Recovery problem of recognizing whether a goal or failure region has been achieved. Based on these methods and strategies for visually-cued camera control, we have built a robot surveillance system in which one mobile robot navigates to a viewing position from which it has an unobstructed view of a goal region, and then uses visual recognition to detect when a specific target has entered the region.

8.2 Products

Our animations and demos are online at http://www.cs.dartmouth.edu/~brd/papers/ICRA2000/.

Books

Algorithmic and Computational Robotics: New Directions, (with K. Lynch and D. Rus) A. K. Peters (Boston: 2001), 408 pp.

Articles and Chapters in Multiply-Authored Books

1. "Constrained Prehensile Manipulation: Distributed Manipulation with Ropes" (with L. Gariepy and D. Rus), in *Distributed Manipulation*, ed. K. Böhringer et al., Kluwer Academic Publishing (2000), pp. 29–48.

2. "Experiments in Constrained Prehensile Manipulation: Distributed Manipulation with Ropes" (with L. Gariepy and D. Rus), in *Experimental Robotics VI*, Lecture Notes in Control and Information Sciences, vol. 250, ed. P. Corke et al., Springer-Verlag, 2000, pp. 25–36.

Invited Papers

1. "Spatial Aggregation in Scientific Data Mining," (with C. Bailey-Kellogg and F. Zhao), First SIAM Conference on Computational Science and Engineering, September 21-24, 2000, Wyndham City Center Hotel, Washington, DC.

2. "Haptics for Animation Motion Control" (with F. Henle), The Fourth PHANTOM Users Group Workshop (PUG99), October 9-12, 1999 (Massachusetts Institute of Technology).

   Received Best Paper Award.

3. "Prototyping Animation Motion Control with Haptic Programmable Force Fields", Workshop on "Motion Support in Virtual Prototyping," (with Fred Henle) Stanford, May 5-7, 1999.

4. "Experiments in Constrained Prehensile Manipulation: Distributed Manipulation with Ropes," (with L. Gariepy and D. Rus) IEEE International Conference on Robotics and Automation, Workshop on Distributed Manipulation (Detroit) May 11, 1999.

Papers in Refereed Journals

1. "Part Orientation with One or Two Stable Equilibria Using Programmable Vector Fields," (with K.-F. Böhringer, F. Lamiraux, and L. Kavraki), IEEE Transactions on Robotics and Automation, 16(2), April 2000, pp. 157-170.

2. "Sensorless Manipulation Algorithms Using a Vibrating Surface," (with K.-F. Böhringer, V. Bhatt, and K. Goldberg) *Algorithmica*, 26**(3/4)**, March/April (2000). Special Issue on Algorithmic Foundations of Robotics, pp. 389–429.

3. "Visibility-Based Planning of Sensor Control Strategies," (with A. J. Briggs), *Algorithmica,* 26**(3/4)**, March/April (2000), pp. 364–388.

4. "Mobile Robot Self-Localization without Explicit Landmarks," (with R. G. Brown), *Algorithmica*, Special Issue on Algorithmic Foundations of Robotics 26**(3/4)**, March/April (2000), pp. 515–559.

Papers in Refereed Conferences

1. "Distributed Manipulation of Multiple Objects Using Ropes," (with L. Gariepy and D. Rus), Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), San Francisco (April, 2000) pp. 450–457.

2. "Practical Mobile Robot Self-Localization," (with J. Howell), Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), San Francisco (April, 2000) pp. 3485–3492.

3. "Experiments in Constrained Prehensile Manipulation: Distributed Manipulation with Ropes," (with L. Gariepy and D. Rus) *International Symposium on Experimental Robotics (ISER)*, Sydney, Australia (1999).

8.3 Contributions

The contributions of this project are discussed above.

9 Graphics and Animation (Bruce Donald)

9.1 Activities and Findings

In our SIGGRAPH'2000 paper, we apply our work on geometric modeling to computer animation. Our goal is to embed free-form constraints into a graphical model. With such constraints a graphic can maintain its visual integrity-and break rules tastefully-while being manipulated by a casual user. A typical parameterized graphic does not meet these needs because its configuration space contains nonsense images in much higher proportion than desirable images, and the casual user is apt to ruin the graphic on any attempt to modify or animate it.

We therefore model the small subset of a given graphic's configuration space that maps to desirable images. In our solution, the basic building block is a simplicial complex-the most practical data structure able to accommodate the variety of topologies that can arise. The configuration-space model can be built from a cross product of such complexes. We describe how to define the mapping from this space to the image space. We show how to invert that mapping, allowing the user to manipulate the image without understanding the structure of the configuration-space model. We also show how to extend the mapping when the original parameterization contains hierarchy, coordinate transformations, and other nonlinearities.

9.2 Products

Patents

System for Image Manipulation and Animation Using Embedded Constraint Graphics (with J. T. Ngo), U.S. Patent #5,933,150, issued August 3, 1999: 91 claims.

Publications

1. "Accessible Animation and Customizable Graphics via Simplicial Configuration Modeling" (with T. Ngo, D. Cutrell, J. Dana, L. Loeb, and S. Zhu), in Proc. ACM SIGGRAPH – the 27^{th} International Conference on Computer Graphics and Interactive Techniques (New Orleans) July, 2000, pp. 403–410.

2. "Using Haptic Vector Fields for Animation Motion Control," (with F. Henle), Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), San Francisco (April, 2000) pp. 3435–3442.

9.3 Contributions

The contributions of this project are discussed above.

10 Computational Geometry (Scot Drysdale)

The following describes the work that I have done that was supported by the departmental network and the systems administration support supplied by the infrastructure grant.

10.1 Activities and Findings

I comment on both research and education activities in this section.

10.1.1 Research Activities

My major research activities in Computational Geometry involved Voronoi Diagrams and various types of triangulations. These include study of the 2-point site Voronoi diagram, work on minimum triangulations, and recent work on minimum-bend path computations. I have been reading papers in the area of surface reconstruction algorithms, but have not made contributions in this area.

The 2-point site Voronoi diagram is a generalization of the standard Voronoi diagram where the distance measures the "distance" from a point to a pair of sites. Some examples are the sum of the distances and the product of the distances from the given point to the pair of sites. G. Barequet, M. Dickerson, and I found algorithms and time and space bounds for a number of these distance functions.

I also studied an algorithm for computing the minimum weight triangulation (MWT) of a set of points invented by Dickerson. The algorithm works surprisingly well in practice, although nobody really understands why. I invented a method for speeding up the algorithm by eliminating possible edges quickly by using an exclusion region around the edge. The edge cannot be in the MWT the two halves of the exclusion region both contain a point from the set. McElfresh, Snoeyink, and I found a larger exclusion region than the ones previously known. We also showed that a related low-weight triangulation has no such exclusion region.

Cliff Stein and I have been supervising David Wagner as a graduate student. Part of his thesis is an algorithm that finds the minimum-bend rectilinear path between a pair of points in the presence of rectilinear obstacles in $O(n^2 logn)$ time. This improves on an $O(n^3)$ algorithm that appeared in a robotics paper.

10.1.2 Educational Activities and Outreach

One educational activity is serving as a member and now as chair of the development committee for the Advanced Placement exam in Computer Science. I helped with the change or the exam language from C++ to Java and in the development of a new Case Study. I helped write a number of AP exams and also made two presentations at SIGCSE and one at an AP Annual Meeting on the change of the exam language from C++ to Java.

A second activity was serving as Principal Lecturer for the DIMACS Reconnect Program (Reconnecting Teaching Faculty to the Mathematical Sciences Research Program), August 11-17, 2002. I gave a week-long series of lectures on the use of Voronoi diagrams and Delaunay triangulations in surface reconstruction algorithms, particularly the Crust algorithm (and its variants) and Co-cone algorithm.

I have also written articles about using examples from Computational Geometry in undergraduate algorithms courses and the place of mathematics in the Computer Science curriculum.

Finally, I have co-authored a Discrete Mathematics textbook that is designed for Computer Science students who have already taken the introductory course and are taking or have taken Data Structures. Most discrete mathematics textbooks do not assume such a background, so are left with a choice: use no CS examples (and thus risk appearing irrelevant to CS students) or teach the computer science needed to understand the CS examples (which is time-consuming and is often attempted by mathematicians who may not know Computer Science). Our book is able to use hashing to motivate the study of probability, recursive sorting and searching algorithms to motivate the study of recurrence relations and their solution, binary search trees to motivate the study of graphs and trees and as an area in which to practice inductive proofs, etc.

10.2 Products

1. M. Dickerson and R. L. Scot Drysdale,"The Undergraduate Algorithms Course and Recent Research in Computational Geometry," *Proceedings of The Consortium for Computing in Small Colleges Third Annual Northeastern Conference*, April 24-25, 1998.

2. M. Dickerson and S. Drysdale,"The undergraduate algorithms course and recent research in computational geometry," *Journal of Computing in Small Colleges (JCSC)* 13:5 (1998) 173-186.

3. B. Chazelle, N. Amenta, T. Asano, G. Barequet, M. Bern, J.-D. Boissonnant, J. Canny, K. Clarkson, D. Dobkin, B. Donald, S. Drysdale, H. Edelsbrunner, D. Eppstein, A. R. Forrest, S. Fortune, K. Goldberg, M. T. Goodrich, L. J. Guibas, P. Hanrahan, C. M. Hoffmann, D. Huttenlocher, H. Iami, D. Kirkpatrick, D.T. Lee, K. Mehlhorn, V. Milenkovic, J. Mitcell, M. Overmars, R. Pollack, R. Seidel, M. Sharir, J. Snoeink, G. T. Toussaint, S. Teller, H. Voelcker, E. Welzl, and C.-K. Yap,"The computational geometry impact task force report," *Advances in Discrete and Computational Geometry*, Contemporary Mathematics, 223, American Mathematical Society, 407-463, 1999.

4. G. Barequet, M. T. Dickerson, and R. L. S. Drysdale,"2-point Site Voronoi Diagrams," *Proc. 6th Workshop on Algorithms and Data Structures*, Vancouver, British Columbia, Canada, Lecture Notes in Computer Science, 1663, Springer-Verlag, 219-230, August 1999.

5. R. L. (Scot) Drysdale, Scott McElfresh, and Jack Scott Snoeyink,"On Exclusion Regions for Optimal Triangulations," *Discrete Applied Mathematics* 109:1-2(2001), 49-65.

6. G. Barequet, M. T. Dickerson, R. L. S. Drysdale, and D. Guertin,"2-point Site Voronoi Diagrams," Videotape and abstract, *Proceedings of the Seventeenth ACM Symposium on Computational Geometry*, June 2001, 323-324.

7. Gill Barequet, Matthew T. Dickerson, and Robert L. Scot Drysdale, "2-point Site Voronoi Diagrams," *Discrete Applied Mathematics*, 122:1-3(2002), 37-54.

8. Scot Drysdale, Judith Hromcik, Mark Allen Weiss, and Reg Hahne,"Java in the Morning Java in the Evening... Java in 2004, " *Proceedings of the 34th SIGCSE Technical Symposium on Computer Science Education*, Feb. 19-22, 2003, 271-272.

9. Kim B. Bruce, Robert L. Scot Drysdale, Charles Keleman, and Allen Tucker, "Why Math?" *Communications of the ACM*, 46:9(Sept. 2003), 40-44.

10. Scot Drysdale, Judith Hromcik, David Reed, and Reg Hahne,"The Year in Review... Changes and Lessons Learned in the Design and Implementation of the AP CS Exam in Java," to appear in *Proceedings of the 35th SIGCSE Technical Symposium on Computer Science Education*, Feb. 23-26, 2005.

11. Kenneth Bogart, Clifford Stein, and Robert L. Drysdale, *Discrete Mathematics for Computer Science*, ISBN 1-930190-86-7, Key College Publishing, in press.

12. R. L. S. Drysdale, C. Stein, and D. Wagner, "The Rectilinear Minimum Bends Distance Problem in Three Dimensions," submitted to the *ACM Symposium on Computational Geometry*.

10.3 Contributions

The research contributions have helped advance the discipline of computational geometry by finding better algorithms to solve various problems.

My work on the AP development committee helped move high school Computer Science classes to an object-oriented approach more in line with current college courses. A number of my activities were aimed at helping high school teachers make that transition.

Teaching in the DIMACS Reconnect program has helped professors in small colleges learn about an interesting new research area that they can explore themselves and can share with their students. It has rekindled an interest in research in some of them.

My hope is that the discrete mathematics textbook will help computer science students learn discrete math by applying it to problems that they care about. I hope that it will have an impact on the way that discrete mathematics is taught.

11 Digital Image Tampering (Hany Farid)

11.1 Activities and Findings

11.1.1 Major Research Activities

One of our main research goals has been the development of computational techniques for the detection of digital tampering. This work has three main directions:

1. detecting the presence of hidden messages (steganography)
2. detecting traces of digital manipulation that change the content/meaning of an image (e.g., the "gluing" together of two photographs)
3. digital art authentication

We are also actively involved in several other areas of research:

1. Virtual Reality and Architecture: we developed computational tools for reconstructing ancient and modern architectural monuments.
2. Medical Imaging: we are working on medical image registration, mapping neural pathways using diffusion tensor MRI, and early detection of disease using fMRI.
3. Neural Binding: we are working to understand the hotly debated topic of binding by temporal synchrony.
4. Landscape Topography: we are developing computational tools for the automatic extraction of landscape topography from photographs.

11.1.2 Major Findings

1. **Steganography:** we have shown that, over a broad range of natural images, certain higher-order statistics in a wavelet decomposition are highly similar. We have also shown that when a message is embedded into an image, in a manner that is imperceptible to the human eye, these statistics are dramatically altered. As such, we are able to detect the presence of covert communication.

2. **Digital Tampering:** we have developed several techniques for determining if an image has been tampered with from the time of its recording. These approaches work by observing that certain forms of tampering leave behind specific statistical correlations that can be quantified and detected.

3. **Computer Graphics or Photographic:** Computer graphics rendering software is capable of generating highly photorealistic images that are often very difficult to differentiate from photographic images. We have, however, developed a method for differentiating between photographic and computer generated (photorealistic) images. Specifically, we have shown that a statistical model based on first- and higher-order wavelet statistics reveals subtle but significant differences between photographic and photorealistic images. This work has begun to have interesting legal applications in light of the 2002 United States Supreme Court ruling that effectively legalized "virtual child pornography".

4. **Art Forgeries:** We have developed a mathematical technique that can classify various parts of a painting as belonging to one or more artists. Beginning with a high-resolution digital scan, our technique works by looking for statistical differences across the canvas. We have also applied this technique to detecting art forgeries – in collaboration with the Metropolitan Museum of Art, we have analyzed drawings by Bruegel and find that we are able to perfectly distinguish between authentic drawings and known forgeries.

5. **Virtual Reality and Architecture:** we have previously digitally reconstructed an ancient Egyptian tomb from a series of photographs. We have applied these techniques to digitally reconstruct the Orozco murals at Dartmouth, Pomona College and New School University. This work was part of a larger Orozco exhibit at the San Diego Museum, allowing visitors to experience, in 3-D, the beauty of these powerful murals. This exhibit was also at Dartmouth's Hood Museum.

6. **Medical Imaging:** we have developed an elegant and powerful computational framework for the registration of medical images. This technique overcomes many of the limitations of existing approaches.

7. **Neural Binding:** we have provided experimental evidence that suggests that a particular model of neural binding, based on temporal synchrony, has a serious flaw. This work is particularly important as expensive and time consuming experiments may be devoted to the search of neural mechanisms, that we believe simply don't exist.

11.1.3 Opportunities for Training and Development

Several undergraduate and graduate students are actively participating in the research described above.

11.1.4 Outreach

I have taught at MathCamp, a program organized by Mathematicians at Harvard and UC Berkeley. This program is geared towards pre-college students and is intended to immerse its participants in deep mathematical thought. I am also developing a summer robot camp to be hosted in the Computer Science Department.

11.2 Publications and Products

Journal

1. A.C. Popescu and H. Farid. Exposing Digital Forgeries in Color Filter Array Interpolated Images. *IEEE Transactions on Signal Processing*, 2005 (in press).
2. S. Lyu and H. Farid. How Realistic is Photorealistic? *IEEE Transactions on Signal Processing*, 53(2):845-850, 2005.
3. A.C. Popescu and H. Farid. Exposing Digital Forgeries by Detecting Traces of Re-sampling. *IEEE Transactions on Signal Processing*, 53(2):758-767, 2005.
4. S. Lyu, D. Rockmore and H. Farid. A Digital Technique for Art Authentication. *Proceedings of the National Academy of Sciences*, 101(49):17006-17010, 2004.
5. H. Sun, D.W. Roberts, H. Farid, Z. Wu, A. Hartov and K.D. Paulsen. Cortical Surface Tracking Using a Stereoscopic Operating Microscope. *Neurosurgery*, 2004 (in press).
6. M.J. Bravo and H. Farid. Search For a Category Target in Clutter. *Perception*, 33:643-652, 2004
7. H. Farid and E.P. Simoncelli. Differentiation of Discrete Multi-Dimensional Signals. *IEEE Transactions on Image Processing*, 13(4):496-508, 2004.
8. M.J. Bravo and H. Farid. Recognizing and Segmenting Objects in Clutter. *Vision Research*, 44(4):385-396, 2004.
9. H. Sun, H. Farid, D.W. Roberts, K. Rick, A. Hartov, and K.D. Paulsen. A Non-Contacting 3-D Digitizer for Use in Image-Guided Neurosurgery. *Steroetactic and Functional Neurosurgery*, 80(1-4):120-124, 2003.
10. R.H. Lilien, H. Farid and B.R. Donald. Probabilistic Disease Classification of Expression-Dependent Proteomic Data from Mass Spectrometry of Human Serum. *Journal of Computational Biology*, 10(6):925-946, 2003.
11. S. Periaswamy and H. Farid. Elastic Registration in the Presence of Intensity Variations. *IEEE Transactions on Medical Imaging*, 22(7):865-874, 2003.
12. M.J. Bravo and H. Farid. Object Segmentation by Top-Down Processes. *Visual Cognition*, 10(4):471-491, 2003.
13. A. Heimsath and H. Farid. Hillslope Topography from Unconstrained Photographs. *Mathematical Geology*, 34(8):929-952, 2002.

14. H. Farid. Temporal Synchrony in Perceptual Grouping: A Critique. *Trends in Cognitive Sciences*, 6(7):284-288, 2002.
15. H. Farid and E.H. Adelson. Synchrony Does Not Promote Grouping in Temporally Structured Displays. *Nature Neuroscience*, 4(9):875-876, 2001.
16. H. Farid and A.C. Popescu. Blind Removal of Lens Distortions. *Journal of the Optical Society of America*, 18(9):2072-2078, 2001.
17. H. Farid. Blind Inverse Gamma Correction. *IEEE Transactions on Image Processing*, 10(10):1428-1433, 2001.
18. M.J. Bravo and H. Farid. Texture Perception on Folded Surfaces. *Perception*, 30(7):819-832, 2001.
19. R. van Ee, B. Anderson, and H. Farid. Occlusion Junctions do not Improve Stereoacuity. *Spatial Vision*, 15(1):45-49, 2001.
20. M.J. Bravo and H. Farid. Effects of 3D Structure on Motion Segmentation. *Vision Research*, 40(6):695-704, 2000.
21. X. Jiang, H. Farid, E. Pistor and R. S. Farid. A New Approach to the Design of Uniquely Folded Thermally Stable Proteins. *Protein Science*, 9:403-416, 2000.
22. E.H. Adelson and H. Farid. Filtering Reveals Form in Temporally Structured Displays. *Science*, 286:2231, 1999.
23. H. Farid and E.H. Adelson. Separating Reflections from Images by use of Independent Components Analysis. *Journal of the Optical Society of America,* 16(9):2136-2145, 1999.

Magazine

24. H. Farid. Is Seeing Believing. *New Scientist*, 179(2411):38-41, Sept. 6, 2003.
25. H. Farid and S. Farid. Unfolding Sennedjem's Tomb. *KMT: A Modern Journal of Ancient Egypt*, 12(1):46-59, 2001.

Refereed Conference Paper

26. J.E. Dobson, J.B. Woodward, S.A. Schwarz, J.C. Marchesini, H. Farid, and S.W. Smith. The Dartmouth Green Grid. *Workshop on High Performance Computing in Academia (in conjunction with International Conference on Computational Science)*, Atlanta, GA, 2005.
27. M.K. Johnson, S. Lyu and H. Farid. Steganalysis in Recorded Speech. *SPIE Symposium on Electronic Imaging*, San Jose, CA, 2005.
28. A.C. Popescu and H. Farid. Statistical Tools for Digital Forensics. *6th International Workshop on Information Hiding*, Toronto, CA, 2004.
29. S. Lyu and H. Farid. Steganalysis Using Color Wavelet Statistics and One-Class Support Vector Machines. *SPIE Symposium on Electronic Imaging*, San Jose, CA, 2004.
30. H. Sun, H. Farid, K. Rick, A. Hartov, D.W. Roberts, and K.D. Paulsen. Estimating Cortical Surface Motion Using Stereopsis for Brain Deformation Models. *Medical Image Computing & Computer Assisted Intervention (MICCAI)*, Montreal, Canada, 2003.
31. J. Ford, H. Farid, F. Makedon, L.A. Flashman, T.W. McAllister, V. Megalooikonomou, and A.J. Saykin. Patient Classification of fMRI Activation Maps. *Medical Image Computing & Computer Assisted Intervention (MICCAI)*, Montreal, Canada, 2003.
32. S. Periaswamy and H. Farid. Elastic Registration with Partial Data. *Second International Workshop on Biomedical Image Registration*, Philadelphia, PA, 2003.
33. H. Farid and S. Lyu. Higher-order Wavelet Statistics and their Application to Digital Forensics. *IEEE Workshop on Statistical Analysis in Computer Vision (in conjunction with CVPR)*, Madison, Wisconsin, 2003.
34. S. Lyu and H. Farid. Detecting Hidden Messages Using Higher-Order Statistics and Support Vector Machines. *5th International Workshop on Information Hiding, Noordwijkerhout*, The Netherlands, 2002.

35. H. Farid. Detecting Hidden Messages Using Higher-Order Statistical Models. *International Conference on Image Processing (ICIP)*, Rochester, NY, 2002.
36. H. Sun, H. Farid, A. Hartov, K.E. Lunn, D.W. Roberts, K.D. Paulsen. Real-time Correction Scheme for Calibration and Implementation of Microscope-based Image-guided Neurosurgery. *SPIE's International Symposium on Medical Imaging*, San Diego, CA, 2002.
37. H. Farid and A.C. Popescu. Blind Removal of Image Non-Linearities. *International Conference on Computer Vision (ICCV)*, Vancouver, Canada, 2001.
38. H. Farid. Reconstructing Ancient Egyptian Tombs. *The International Symposium on Virtual and Augmented Architecture*, Dublin, Ireland, 2001.
39. S. Periaswamy, J.B. Weaver, D.M. Healy Jr., D. Rockmore, P.J. Kostelec, and H. Farid. Differential Affine Motion Estimation for Medical Image Registration. *SPIE's 45th Annual Meeting*, San Diego, CA, 2000.
40. H. Farid and E.H. Adelson. Separating Reflections and Lighting in Images Using Independent Components Analysis. *Computer Vision and Pattern Recognition (CVPR),* June 1999.

Conference Abstract

41. M.J. Bravo and H. Farid. The Depth of Distractor Processing in Search Through Clutter. *Vision Sciences*, Sarasota, FL, 2005.
42. M.J. Bravo and H. Farid. Still Searching a Cluttered Scene. *Vision Sciences*, Sarasota, FL, 2004.
43. V. Maljkovic, P. Martini and H. Farid. The Time-Course of Categorization of Real-Life Scenes with Affective Content. *Vision Sciences*, Sarasota, FL, 2004.
44. H. Sun, H. Farid D. Roberts, K. Rick, A. Kartov, and K. Paulsen. A Non-contacting 3-D Digitizer For Use in Image-Guided Neurosurgery. *American Society for Stereotactic and Functional Neurosurgery*, New York City, 2003.
45. M.J. Bravo and H. Farid. Searching a Cluttered Scene. *Vision Sciences*, Sarasota, FL, 2003.
46. A.M. Heimsath and H. Farid. Hillslope Topography from Unconstrained Photographs. *Transactions of the American Geophysical Union*, 2002.
47. H Farid and E.H. Adelson. Energy versus Synchrony in Perceptual Grouping. *Vision Sciences*, Sarasota, FL, 2002.
48. M.J. Bravo and H. Farid. Segmentation in Clutter. *Vision Sciences*, Sarasota, FL, 2002.
49. S. Inati, H. Farid, K. Sherwin, and S. Grafton. A Global Probabilistic Approach to Fiber Tractography with Diffusion Tensor MRI. *Human Brain Mapping*, Brighton, UK, 2001.
50. M.J. Bravo and H. Farid. Top-Down and Bottom-Up Processes for Object Segmentation. *Vision Sciences*, Sarasota, FL, 2001.
51. J.B. Weaver, S. Periaswamy, H. Farid, D.N. Rockmore, C.J. Kasales, W. Black, and D.M. Healy Jr. Lesion Size Estimation Using Warped Registration of Interval Images. *International Society for Magnetic Resonance in Medicine*, 2001.
52. H. Farid and E.H. Adelson. Standard Mechanisms Can Explain Grouping in Temporally Synchronous Displays. *Investigative Opthalmology and Visual Science*, Fort Lauderdale, FL, 2000.
53. M.J. Bravo and H. Farid. The Role of Object Recognition in Scene Segmentation. *Investigative Opthalmology and Visual Science*, Fort Lauderdale, FL, 2000.
54. M.J. Bravo and H. Farid. Segmentation in 3D. *Investigative Opthalmology and Visual Science,* Fort Lauderdale, FL, 1999.

11.2.1 Books/non-Periodicals

None.

11.2.2 Web Site

www.cs.dartmouth.edu/farid

11.2.3 Other Specific Products

I make freely available an extensive tutorial on the fundamentals of image processing, and a MatLab toolbox for image processing:

www.cs.dartmouth.edu/farid/tutorials

11.3 Contributions

Our work on digital tampering is having immediate applications to a variety of local and federal law enforcement agencies, and is helping the legal community contend with the ever-growing presence of digital media. Our work on digital art authentication is ground-breaking and will bring to the Art world new and exciting tools for the quantitative process of authentication.

In a larger sense, virtually every discipline is involved in the analysis of digital media. I have been developing a curriculum, software, tutorials and active collaborations to help my colleagues and students to meet these challenges.

12 Programming Languages (Chris Hawblitzel)

Our work focused on type-safe programming languages.

12.1 Activities and Findings

We discuss two projects, a low-level, type-safe language called "Clay" and an exploration of safe "remote pointers" to support a process analogy in type-safe languages.

Clay. Modern computers rely on the correctness and security of low-level systems, such as garbage collectors, device drivers, embedded system code. Often, a small amount of low-level code (say, in a firewall, a network interface device driver, or secure coprocessor) is all that stands between a computer system and a hacker trying to break into the system. Even in the absence of malicious outsiders, buggy device drivers often cause annoying system crashes. Given their role as the foundation for higher-level services, one might expect low-level systems to benefit from the static and run-time checks provided by type-safe languages. However, systems programmers usually lean towards assembly language, C, or C++ rather than Java, ML, or Haskell, because they need efficient low-level control of memory. Java's array bounds checking, for example, prevents the buffer overflow attacks that so many C programs are susceptible to, but often imposes extra run-time overhead. Automatic garbage collection prevents dangling pointers, but implementing a garbage collector requires the ability to explicitly free heap objects, a privilege not usually granted to safe language programs. We have implemented a low-level, type-safe language called "Clay" that is powerful enough to express programs that could previously only be written in unsafe languages. It includes sophisticated type system support for safe memory deallocation and array bounds check elimination.

We demonstrated Clay's utility by writing an efficient Cheney-queue copying collector and a mark-sweep collector in the language, and we used the insights from these collectors to develop a new type-theoretic foundation for regions, based on linear memory types. We also used Clay to write safe operating system code that interacts directly with low-level entities, such as devices and interrupts. For example, our ethernet driver reads and writes directly to the registers on an ethernet card, but Clay's type system prevents illegal operations, such as writing to a non-existent register, writing to a register without first establishing the correct register window, accessing the device without holding the proper lock, or overflowing a queue.

Remote pointers. Traditional operating systems use virtual memory to protect the operating system from user processes (and to protect user processes from each other). Many recent systems use safe languages as an alternative protection mechanism; the Java language, for example, protects browsers from applets and servers from servlets. Although safe languages promise many advantages over traditional protection mechanisms, they lack a compelling analogue to traditional operating system processes, which makes safe language systems more difficult to manage than traditional systems. We have developed a process abstraction suitable for safe languages, allowing a host to limit resource each process's usage, revoke access to process's resources, and shut down errant processes [HvE02].

We found that in order to shut down safe language "processes" cleanly, there needs to be a clear boundary between the objects, code, and threads of different processes. This boundary is difficult to draw if processes can share arbitrary objects, so we introduced a new type into the language's type system: "remote pointers" have different types from local pointers, and processes can only share objects through remote pointers. The remote pointer types allow the run-time system to insert revocation checks and thread switches in the appropriate places.

12.2 Products

Software. We distribute the source code for Clay.[3]

[3] http://www.cs.dartmouth.edu/~hawblitz/publish/clay-0.03.tgz

Publications

[HHW04] Chris Hawblitzel, Heng Huang, and Lea Wittie. Composing a well-typed region. Technical Report TR2004-521, Dartmouth College, October 19, 2004.

[HvE02] Chris Hawblitzel and Thorsten von Eicken. Luna: a flexible Java protection system. In *Operating Systems Design and Implementation (OSDI)*, pages 391–403. ACM Press, 2002.

[HWH+04] Chris Hawblitzel, Edward Wei, Heng Huang, Eric Krupski, and Lea Wittie. Low-level linear memory management. In *Workshop on Semantics, Program Analysis, and Computing Environments For Memory Management*, 2004.

[Wit04] Lea Wittie. *Type-Safe Operating System Abstractions*. PhD thesis, Dartmouth College, 2004. Available as Technical Report TR2004-526.

12.3 Contributions

With Clay we showed that safe languages are applicable to a wider class of applications than previously thought, and that these applications can benefit from the strong guarantees that a safe language's type system provides.

Our other project produced a middle ground between traditional operating systems and safe language systems. It allows more flexible, fine-grained sharing between processes than traditional operating systems, but with clearer inter-process boundaries than conventional safe language systems. Our work was one influence on the Java isolation API.

13 Distributed Algorithms (Prasad Jayanti)

Our research on distributed algorithms focused on the design of efficient lock-free (i.e., wait-free or nonblocking) as well as locking protocols.

13.1 Activities and Findings

Our research on wait-free synchronization and locking protocols has produced the following algorithms:

1. A time optimal wait-free algorithm for the wellknown multiwriter snapshot problem [Jay05]
2. An efficient wait-free algorithm for implementing a multiword LL/SC object [JP05]
3. An efficient wait-free algorithm for computing tree functions in time proportional to the depth of the tree [Jay02]
4. Constant time wait-free implementation of LL/SC instructions from compare&swap [JP03]
5. Efficient design of FCFS *abortable* locks [Jay03]
6. Design of FCFS group mutual exclusion locks [JPT03]
7. Transformation of mutual exclusion algorithms into *fast* mutual exclusion algorithms [JPN05]

13.2 Products

Our research has led to four publications at PODC [Jay02, Jay03, JPT03, JP03], and one publication each at SOFSEM, ICDCS and STOC [JPN05, JP05, Jay05].

Publications

[Jay02] P. Jayanti. f-arrays: implementation and applications. In *Proceedings of the 21st Annual Symposium on Principles of Distributed Computing*, pages 270 – 279, 2002.

[Jay03] P. Jayanti. Adaptive and efficient abortable mutual exclusion. In *Proceedings of the 22nd ACM Symposium on Principles of Distributed Computing*, July 2003.

[Jay05] P. Jayanti. An optimal multiwriter snapshot algorithm. In *Proceedings of the 37th ACM Symposium on Theory of Computing (STOC)*, May 2005.

[JP03] P. Jayanti and S. Petrovic. Efficient and practical constructions of ll/sc variables. In *Proceedings of the 22nd ACM Symposium on Principles of Distributed Computing*, July 2003.

[JP05] P. Jayanti and S. Petrovic. Efficient wait-free implementation of multiword ll/sc variables. In *Proceedings of the 25th International Conference on Distributed Computing Systems (ICDCS)*, June 2005.

[JPN05] P. Jayanti, S. Petrovic, and N. Narula. Read/write based fast path transformation for fcfs mutual exclusion. In *SOFSEM 2005: Theory and Practice of Computer Science, 31st Conference on Current Trends in Theory and Practice of Computer Science, LNCS 3381*, January 2005.

[JPT03] P. Jayanti, S. Petrovic, and K. Tan. Fair group mutual exclusion. In *Proceedings of the 22nd ACM Symposium on Principles of Distributed Computing*, July 2003.

13.3 Contributions

Two Ph.D students (King Tan and Srdjan Petrovic) and three undergraduate students (Neha Narula, Rachel Ringel and Sam Slee) worked on the projects as part of their thesis research. King Tan got his Ph.D in 2003 and Srdjan Petrovic is expected to complete the Ph.D in Summer 2005.

14 Mobile Agents (David Kotz, Daniela Rus)

Our group, known as the *D'Agents* project, was funded by DARPA's Control of Agent-Based Systems (CoABS) project during the same period (1998–2004) as this NSF award. We used this funding, and the departmental infrastructure built with this NSF award, to delve ever more deeply into the fundamental challenges facing mobile-agent systems, including related topics in mobile ad hoc wireless networks, sensor networks, market-based control mechanisms for mobile agent systems, information retrieval (a common application domain for mobile agents), and middleware for sensor information processing.

We collaborated closely Professor George Cybenko and Research Engineer Bob Gray at Dartmouth's Thayer School of Engineering, research teams from Lockheed-Martin, the University of Western Florida, and Boeing, as well as non-CoABS researchers from the University of Illinois at Urbana-Champaign. Indeed, we led the Mobility TIE, a deep and productive collaboration that led to specifications for mobility support that became part of the CoABS Grid, and to two lines of research papers: one based on joint scalability experiments, and one based on joint development of an interoperability layer. The Mobility TIE also was involved in a Fleet Battle Experiment (obtaining experimental data that was incorporated into one of our papers [KCG+02]) and the CoAX coalition-forces demonstration.

The result was an extremely productive research team, producing numerous papers, training over two dozen students in Computer Science and in Computer Engineering, incorporating our research in undergraduate and graduate courses, demonstrating working prototypes to key players in government and industry, and laying the intellectual groundwork for several exciting new research efforts now underway.

14.1 Activities and Findings

A mobile agent is a running program that can migrate from host to host at times and to places of its own choosing. A software developer can use the ability to write a migratory program to develop software systems that are more efficient and more robust, particularly in a distributed system that includes wireless networks with low bandwidth and high disconnectivity.

In this project we set out to gain a better understanding of mobile-agent systems, and in particular their scalability and ability to interoperate with other mobile-agent systems. Using models, simulation, and large-scale experiments, and collaborating with the developers of other mobile-agent systems, we performed the first and (to date) most extensive scalability analysis of mobile-agent software. We compared their performance on information-retrieval tasks common to many application domains. We explored their performance benefits in a wireless-network environment. We developed technology that allows a mobile agent written for one mobile-agent system to migrate to and execute in a different mobile-agent system. We also expanded our prior work on the security of mobile agents.

In any mobile-agent system it is important to control the amount of resources consumed by a mobile agent. We used the concept of market-based control to allow hosts to "sell" computation time to visiting agents, setting the price through game-theoretic auction mechanisms. The agents, given a budget and a sequence of tasks to complete, seek low-price hosts that allow them to complete their tasks within budget. The result is a natural, distributed load-balancing scheme that works for self-interested agents and hosts.

Along the way, we delved into related topics in ad hoc wireless networks, sensor networks (see Section 25), information retrieval, and middleware for context-aware computing (see Section 18) and sensor-information processing. We aided the research community through our leadership of conferences and our founding of the Dartmouth Workshop on Transportable Agents. Building on our experience in the field, we also wrote and presented forward-looking articles on mobile-agent research.

We created and delivered at least one course on Mobile Agents and related technologies.

14.1.1 Mobile agents.

The Dartmouth Agents, or D'Agents, system is a mobile-agent system that is distinguished by its support for strong mobility (so that an agent can migrate to a new machine at any point during its execution, implicitly carrying its state), its support for multiple agent languages (so that the agent programmer can select an appropriate language for the application), and its high performance (so that mobile-agent performance can be compared in a meaningful way with the performance of traditional client/server solutions). Although work on D'Agents began in 1994 as part of Gray's Ph.D thesis, we improved the performance, security and fault tolerance of D'Agents under the CoABS program [GCKR02, GCKR00, GCK+02, GKCR98], considered the future of mobile agents and mobile code [KGR02a, KGR02b, KG99b, KG99a], adapted a large existing application of D'Agents for use in the CoAX experiment; and used D'Agents as the starting point for all of our higher-level agent work.

This higher-level work can be divided into four categories. First, one weakness of existing mobile-agent systems was that they do not interoperate. An agent in one agent system can not communicate with the agents of or migrate to an agent platform of another agent system. Dartmouth, in cooperation with CoABS participants from Lockheed Martin and the University of Western Florida, developed the Grid Mobile-Agent System (GMAS), a set of standardized interfaces that allowed an agent conforming to that interface to function properly inside multiple agent systems [GGK+02, GGK+01]. GMAS was incorporated into the CoABS Grid software package.

Second, it is important to understand how the performance of mobile agents compares with that of traditional client/server solutions so that a developer can select the most appropriate implementation strategy for their application. Dartmouth, independently and in cooperation with the same set of partners, undertook a series of experiments that compared mobile-agent and client-server performance for information-processing applications (i.e., applications where information must be retrieved from multiple remote sites and then correlated for human use) [GCK+02, GKP+01a, GKP+01b, KCG+02, KJG+00a, KJG+00b, KCG+00]. Although many of the results conform to intuition (e.g., mobile agents are a better choice when available network bandwidth is low), the analysis of quantitative experimental results filled an important gap in available mobile-agent performance studies. As a companion effort, Dartmouth also examined the difficulties of simulating mobile-agent performance accurately enough to make performance predictions, and developed an accurate simulation package for information-processing applications [Dub04].

Third, many mobile agents have some choice as to which remote hosts to visit and in which order. Scheduling mobile-agent migration in a way that minimizes bandwidth and other resource use, therefore, is both feasible and attractive. Dartmouth considered several variations of the scheduling problem, and developed several algorithms for scheduling parallel agent-based access to remote resources [RSX01, XRS01].

Overall, there were 19 papers in this category [Dub04, GCKR02, GCKR00, GKP+01a, GKP+01b, GKCR98, GCK+02, GGK+02, GGK+01, KGR02a, KGR02b, KG99b, KG99a, KCG+02, KJG+00a, KJG+00b, KCG+00, RSX01, XRS01].

14.1.2 Market-based resource control.

We studied the use of markets to distributively allocate computational resources among software agents in computational systems. Markets proved useful in providing incentives to participate in distributed applications, prioritizing tasks, and adding additional fault tolerance. Additionally, we derived structures to facilitate planning and control the degree of private information disclosure, and structures that allow agents to exchange volatility with performance.

Incentives. As distributed systems get larger, the number of principals that contribute to an application increases and a particular resource owner may not immediately realize benefit from system participation. By charging resource usage to a software agent's account, a resource owner can extract value from the use of its wares; value that could be exchanged for services at other sites or for another, possibly legal tender, resource [BKR98a, BKR99a].

Since each agent has a limited endowment with which to acquire resources, an agent's potential to wreak havoc on the network is limited. A long-lived agent must either choose to act at times with low resource-usage demand, or receive a large endowment from its owner. The budget constraint provides a degree of fault tolerance against malicious and buggy code [BKR97, BKR98a, BKR99a, Bre01a].

Finally, price systems allow flexible discrimination among agents. More well-endowed agents compute more quickly than less-endowed agents, especially during times of resource contention. Our experiments show that systems using our resource allocation can effectively serve wealthy agents, even when total load far exceeds capacity [BKR99c, BMcI+99, BMcI+00, BKR+01b, Bre01a, BMcI+03].

Structures. Providing market structure only partially solves resource-allocation problems; an agent must be able to interact with the market and plan to complete its goals. Towards this end, we studied many different auction schemes in which agents bid for computational resources [BKR98b, BKR99c, BMcI+99, BMcI+00, BKR+01b, BKR01a, Bre01a]. We found a tradeoff in the information regarding preferences agents disclosed and efficiency of the market and in the complexity required to calculate prices [BKR01a, Bre01a].

Most of the market mechanisms we studied were demand driven. An agent purchased resources immediately prior to use. To alleviate risk, however, we instrumented call options in our market where agents could purchase a contract to guarantee computation at a fixed price in the future. We extended the Cox, Ross, Rubinstein option-pricing model to allow agents to integrate demand and reservation-based computing according to their preference towards risk [BKR00, Bre01a].

Each of the planning algorithms we implemented relied on estimates of an agent's planned consumption. In real-life applications, resource use is difficult to forecast, so we tested our algorithms using flawed estimates to find that our algorithms adapted an agent's plans as the estimation errors became more apparent [Bre01a, BMcI+03]. Finally, we implemented our resource-allocation and planning algorithms as part of a Linux real-time kernel scheduling process to regulate mobile processes [CG00].

There are 17 papers in this category [BKR98b, BKR98c, BMcI+00, BKR+01b, BMcI+99, BKR01a, BMcI+03, BKR99b, BKR99c, BKR98a, BKR97, BKR99a, BKR00, Bre01a, Bre01b, CG00].

14.1.3 Conclusions

We came to many conclusions; the details are in our papers.

- Mobile code, and often mobile agents, are an efficient mechanism for information filtering and retrieval applications. We learned that a carefully constructed implementation can be scalable and can outperform a traditional remote procedure call approach, in situations where the communication bandwidth is limited and the remote execution of application logic can reduce the communication needed. We conducted the largest scalability experiments of any mobile-agent system, and indeed directly compared the scalability three mobile-agent systems.

- Market-based control is an effective method for balancing load in a distributed system, with self-interested mobile agents seeking out the lowest price to accomplish their tasks. We devised auction-based schemes to support this idea.

- Mobile agents can be used to carry messages through an ad hoc wireless network that is prone to disconnection. Indeed, the agent can ask the network nodes to move (physically) so that the message can reach a currently unreachable portion of the network. We demonstrated algorithms that make this possible.

- The ideas behind mobile agents can be generalized to our Solar framework that can support distributed information filtering, processing, and dissemination. Here, mobile code allows the computation to move closer to the data, can be shared to allow for scalability in the face of large numbers of applications desiring the same information stream, and can be rearranged to balance load.

- And many others.

14.2 Products

The D'Agents web site provides access to the software and publications from this project.[4]

Our research resulted in 34 papers or reports, numerous demonstrations of research prototypes at government and industry labs, transfer of our technology to other research efforts (including contributions to the CoABS Grid software base), dozens of research students trained, and extensive contributions to both the research and government communities. In this report we list only those papers that were produced with the involvement of faculty or students in the Department of Computer Science.

We published 7 journal papers [BMcI+03, GCK+02, KGR02a, KCG+02, LR03, LPDR03, ALR03].

We presented 11 papers at refereed conferences and workshops [BMcI+00, BKR01a, BKR98a, BKR99a, BKR00, GKP+01a, GGK+02, KG99b, KJG+00a, RSX01, XRS01].

We released 14 technical reports [BKR98c, BMcI+99, BKR99c, BKR97, Bre01b, CG00, Dub04, GCKR00, GKP+01b, GGK+01, KGR02b, KJG+00b, KCG+00, MK02].

We contributed 4 book chapters [BKR+01b, BGM+99, GCKR02, GKCR98].

Finally, we wrote 3 other unrefereed or unpublished papers [BKR98b, BKR99b, KG99a].

Furthermore, one student completed a Ph.D thesis in the project [Bre01a], and 2 undergraduates completed a Senior Honors thesis [CG00, Dub04].

Since most of the technical reports were ultimately published as conference or journal papers, and some of the conference papers were expanded and published as journal papers, there are not 34 unique papers represented above; nonetheless, there are 27 peer-reviewed publications.

Publications

[BGM+99] Brian Brewington, Robert Gray, Katsuhiro Moizumi, David Kotz, George Cybenko, and Daniela Rus. Mobile agents for distributed information retrieval. In Matthias Klusch, editor, *Intelligent Information Agents*, chapter 15, pages 355–395. Springer-Verlag, 1999.

[BKR97] Jonathan Bredin, David Kotz, and Daniela Rus. Market-based resource control for mobile agents. Technical Report PCS-TR97-326, Dept. of Computer Science, Dartmouth College, December 1997.

[BKR98a] Jonathan Bredin, David Kotz, and Daniela Rus. Market-based resource control for mobile agents. In *Proceedings of the Second International Conference on Autonomous Agents*, pages 197–204. ACM Press, May 1998.

[BKR98b] Jonathan Bredin, David Kotz, and Daniela Rus. Utility driven mobile-agent scheduling. Unpublished, October 1998.

[BKR98c] Jonathan Bredin, David Kotz, and Daniela Rus. Utility driven mobile-agent scheduling. Technical Report PCS-TR98-331, Dept. of Computer Science, Dartmouth College, May 1998. Revised October 3, 1998.

[BKR99a] Jonathan Bredin, David Kotz, and Daniela Rus. Economic markets as a means of open mobile-agent systems. In *Proceedings of the Workshop "Mobile Agents in the Context of Competition and Cooperation (MAC3)" at Autonomous Agents '99*, pages 43–49, May 1999.

[BKR99b] Jonathan Bredin, David Kotz, and Daniela Rus. Mobile-agent planning in a market-oriented environment. Accepted at, and withdrawn from, ASA/MA '99, August 1999.

[4] http://agent.cs.dartmouth.edu

[BKR99c] Jonathan Bredin, David Kotz, and Daniela Rus. Mobile-agent planning in a market-oriented environment. Technical Report PCS-TR99-345, Dept. of Computer Science, Dartmouth College, May 1999. Revision 1 of May 20, 1999.

[BKR00] Jonathan Bredin, David Kotz, and Daniela Rus. Trading risk in mobile-agent computational markets. In *Persented at the Sixth International Conference on Computing in Economics and Finance*, Barcelona, Spain, July 2000. No proceedings available.

[BKR01a] Jonathan Bredin, David Kotz, and Daniela Rus. The role of information in computational-resource allocation, for the TASK electronic commerce REF. Invited paper at the DARPA TASK PI meeting, May 2001.

[BKR+01b] Jonathan Bredin, David Kotz, Daniela Rus, Rajiv T. Maheswaran, Çagri Imer, and Tamer Başar. A market-based model for resource allocation in agent systems. In Franco Zambonelli, editor, *Coordination of Internet Agents Models, Technologies, and Applications*, chapter 17, pages 426–441. Springer-Verlag, 2001.

[BMcI+99] Jonathan Bredin, Rajiv T. Maheswaran, Çagri Imer, Tamer Başar, David Kotz, and Daniela Rus. A game-theoretic formulation of multi-agent resource allocation. Technical Report PCS-TR99-360, Dept. of Computer Science, Dartmouth College, October 1999.

[BMcI+00] Jonathan Bredin, Rajiv T. Maheswaran, Çagri Imer, Tamer Başar, David Kotz, and Daniela Rus. A game-theoretic formulation of multi-agent resource allocation. In *Proceedings of the Fourth International Conference on Autonomous Agents*, pages 349–356. ACM Press, June 2000.

[BMcI+03] Jonathan Bredin, Rajiv T. Maheswaran, Çagri Imer, Tamer Başar, David Kotz, and Daniela Rus. Computational markets to regulate mobile-agent systems. *Autonomous Agents and Multi-Agent Systems*, 6(3):235–263, May 2003.

[Bre01a] Jonathan L. Bredin. *Market-based Control of Mobile-agent Systems*. PhD thesis, Dept. of Computer Science, Dartmouth College, June 2001. Available as Dartmouth Computer Science Technical Report TR2001-408.

[Bre01b] Jonathan L. Bredin. Market-based control of mobile-agent systems. Technical Report TR2001-408, Dept. of Computer Science, Dartmouth College, June 2001. Ph.D Dissertation.

[CG00] Ezra E. K. Cooper and Robert S. Gray. An economic CPU-time market for D'Agents. Technical Report TR2000–375, Dept. of Computer Science, Dartmouth College, June 2000.

[Dub04] Nikita E. Dubrovsky. Mobile agents simulation with DaSSF. Technical Report TR2004-499, Dept. of Computer Science, Dartmouth College, June 2004.

[GCK+02] Robert S. Gray, George Cybenko, David Kotz, Ronald A. Peterson, and Daniela Rus. D'Agents: Applications and performance of a mobile-agent system. *Software— Practice and Experience*, 32(6):543–573, May 2002.

[GCKR00] Robert S. Gray, George Cybenko, David Kotz, and Daniela Rus. Mobile agents: Motivations and state of the art. Technical Report TR2000-365, Dept. of Computer Science, Dartmouth College, 2000.

[GCKR02] Robert S. Gray, George Cybenko, David Kotz, and Daniela Rus. Mobile agents: Motivations and state of the art. In Jeffrey Bradshaw, editor, *Handbook of Agent Technology*. AAAI/MIT Press, 2002. Accepted for publication. Draft available as Technical Report TR2000-365, Department of Computer Science, Dartmouth College.

[GGK+01] Arne Grimstrup, Robert Gray, David Kotz, Thomas Cowin, Greg Hill, Niranjan Suri, Daria Chacón, and Martin Hofmann. Write once, move anywhere: Toward dynamic interoperability of mobile agent systems. Technical Report TR2001-411, Dept. of Computer Science, Dartmouth College, July 2001.

[GGK+02] Arne Grimstrup, Robert Gray, David Kotz, Maggie Breedy, Marco Carvalho, Thomas Cowin, Daria Chacón, Joyce Barton, Chris Garrett, and Martin Hofmann. Toward dynamic interoperability of mobile agent systems. In *Proceedings of the Sixth IEEE International Conference on Mobile Agents*, volume 2535 of *Lecture Notes in Computer Science*, pages 106–120, October 2002.

[GKCR98] Robert S. Gray, David Kotz, George Cybenko, and Daniela Rus. D'Agents: Security in a multiple-language, mobile-agent system. In Giovanni Vigna, editor, *Mobile Agents and Security*, volume 1419 of *Lecture Notes in Computer Science*, pages 154–187. Springer-Verlag, 1998.

[GKP+01a] Robert S. Gray, David Kotz, Ronald A. Peterson, Jr., Joyce Barton, Daria Chacón, Peter Gerken, Martin Hofmann, Jeffrey Bradshaw, Maggie Breedy, Renia Jeffers, and Niranjan Suri. Mobile-agent versus client/server performance: Scalability in an information-retrieval task. In *Proceedings of the Fifth IEEE International Conference on Mobile Agents*, volume 2240 of *Lecture Notes in Computer Science*, pages 229–243, Atlanta, Georgia, December 2001. Springer-Verlag. A corrected version of this paper is available on the Dartmouth web site.

[GKP+01b] Robert S. Gray, David Kotz, Ronald A. Peterson, Jr., Peter Gerken, Martin Hofmann, Daria Chacón, Greg Hill, and Niranjan Suri. Mobile-agent versus client/server performance: Scalability in an information-retrieval task. Technical Report TR2001-386, Dept. of Computer Science, Dartmouth College, January 2001.

[KCG+00] David Kotz, George Cybenko, Robert S. Gray, Guofei Jiang, Ronald A. Peterson, Martin O. Hofmann, Daria A. Chacon, Kenneth R. Whitebread, and James Hendler. Performance analysis of mobile agents for filtering data streams on wireless networks. Technical Report TR2000-377, Dept. of Computer Science, Dartmouth College, October 2000.

[KCG+02] David Kotz, George Cybenko, Robert S. Gray, Guofei Jiang, Ronald A. Peterson, Martin O. Hofmann, Daria A. Chacón, Kenneth R. Whitebread, and James Hendler. Performance analysis of mobile agents for filtering data streams on wireless networks. *Mobile Networks and Applications*, 7(2):163–174, April 2002.

[KG99a] David Kotz and Robert S. Gray. Mobile agents and the future of the Internet. *ACM Operating Systems Review*, 33(3):7–13, August 1999.

[KG99b] David Kotz and Robert S. Gray. Mobile code: The future of the Internet. In *Proceedings of the Workshop "Mobile Agents in the Context of Competition and Cooperation (MAC3)" at Autonomous Agents '99*, pages 6–12, May 1999.

[KGR02a] David Kotz, Robert Gray, and Daniela Rus. Future directions for mobile-agent research. *IEEE Distributed Systems Online*, 3(8), August 2002. Based on a conversation with Jeff Bradshaw, Colin Harrison, Guenter Karjoth, Amy Murphy, Gian Pietro Picco, M. Ranganathan, Niranjan Suri, and Christian Tschudin.

[KGR02b] David Kotz, Robert Gray, and Daniela Rus. Future directions for mobile-agent research. Technical Report TR2002-415, Dept. of Computer Science, Dartmouth College, January 2002. Based on a conversation with Jeff Bradshaw, Colin Harrison, Guenter Karjoth, Amy Murphy, Gian Pietro Picco, M. Ranganathan, Niranjan Suri, and Christian Tschudin.

[KJG+00a] David Kotz, Guofei Jiang, Robert Gray, George Cybenko, and Ronald A. Peterson. Performance analysis of mobile agents for filtering data streams on wireless networks. In *Proceedings of the Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM)*, pages 85–94. ACM Press, August 2000.

[KJG+00b] David Kotz, Guofei Jiang, Robert Gray, George Cybenko, and Ronald A. Peterson. Performance analysis of mobile agents for filtering data streams on wireless networks. Technical Report TR2000-366, Dept. of Computer Science, Dartmouth College, May 2000.

[RSX01] Daniela Rus, Clifford Stein, and Rong Xie. Scheduling multi-task multi-agent systems. In *Proceedings of the Fifth International Conference on Autonomous Agents*, pages 159–160. ACM Press, 2001. Poster abstract.

[XRS01] Rong Xie, Daniela Rus, and Cliff Stein. Scheduling multi-task agents. In *Proceedings of the Fifth IEEE International Conference on Mobile Agents*, volume 2240 of *Lecture Notes in Computer Science*, pages 260–276, Atlanta, Georgia, December 2001. Springer-Verlag.

Software. We updated and released the source code for our D'Agents mobile-agent system. We have reports of this software being used for research in other universities and corporate labs.

As part of the CoABS Grid effort, we co-developed the Grid Mobile Agent System (GMAS) with Lockheed Martin's ATL and the University of Western Florida's IHMC. GMAS was a major extension to the Grid that defined a Java API for mobile Grid agents. Through proxy-based launchers and bridges, a Grid agent could launch a mobile agent into any Java-based mobile-agent system that supported this API, and could communicate with the mobile agent or any other agent on that system. The mobile agent also could migrate *between* different agent systems, as long as all of the systems supported the API. As a test of GMAS functionality, we added support for the GMAS API to D'Agents, EMAA, and NOMADS, the mobile-agent systems of Dartmouth, Lockheed Martin, and UWF respectively, and demonstrated an application agent migrating sequentially through the three different agent platforms. This same agent was used to provide medical-monitoring support in the CoAX effort, a joint experiment and technology demonstration designed to explore the efficacy of the CoABS software in coalition applications (see below for more information). With GMAS integrated into the main CoABS Grid distribution, Grid applications easily can take advantage of the bandwidth and latency advantages of mobile agents.

14.3 Contributions

We contributed substantially to a better understanding of mobile-agent systems and related areas of ad hoc wireless networks, sensor networks, information retrieval, and middleware for context-aware computing and sensor-information processing.

In addition to contributing numerous papers to the computer science and engineering communities interested in mobile agents and mobile code, we led the community by creating workshops to build the community, and as general chair or program chairs of all the major conferences in the field. Building on our experience in the field we also wrote and presented forward-looking articles on mobile-agent research.

Finally, we contributed ideas and technology to serve in the defense of the nation.

Technology transfer. Our CoABS efforts had a significant impact on our other DoD-funded research efforts, and we were able to use our results and prototypes to support that research.

As one example, our mobile-agent technology was used in an application for medical monitoring of victims on a battlefield, funded by the U.S. Army CECOM, and a related application for medical monitoring of astronauts in long-duration space flight, funded by the NASA Institute for Advanced Concepts. The Grid was used to provide interoperability between D'Agents and other agent systems, and was used for the Guardian Angel

system developed by Lockheed Martin. This project was led by Dr. Sue McGrath, a collaborator who learned of our technology while at Lockheed Martin ATL and then joined Dartmouth College to pursue the research here.

We were deeply involved with the CoAX TIE, an international research effort to explore the use of agent technology to support coalition forces. Bob Gray and Susan McGrath provided a version of the CECOM medical-monitoring system for use in this TIE. This system uses our Grid Mobile Agent System (GMAS) to send monitoring agents to the location(s) of injured soldiers and sailors, allowing medics from one country to efficiently monitor wounded sailors from another country. We demonstrated the final version, with a (simple) graphical user interface for the medics, as part of the final CoAX presentations in October 2002. In the CoAX presentation, the system was used to monitor the health of casualties on board an Australian ship, *HMAS Coonawarra*, which was damaged during a submarine attack. Representatives from the Marine Corps expressed significant interest in the system during the Tech Fair following the presentations; McGrath and Gray are following up with these representatives.

We also made several efforts to raise awareness of our results and technologies within the DoD world and the corporate world, above and beyond the impact of our publications, presentations at conferences, and presentations at university, corporate, and government labs. In particular, we sent a high-level summary of our research, with pointers to detailed results, to a hundred or so key players in the DoD world.

On many occasions we discussed transferring our technology...

- with BAE Systems (British AeroSpace) about using Dartmouth's agent and security work in their advanced communications systems.
- with the Director of Research at the FAA about transitioning some of our work to that setting.
- with Jeff Hughes of Wright Patterson AFB on a potential OSD Secure Applications Initiative.
- with the Grid Forum.[5] Kotz participated in the first meeting of the Grid Forum, a large nationwide group that is trying to coordinate the interconnection of several metacomputing efforts, from NASA, NSF, DOE, DOD, and numerous others. Kotz was there to investigate potential technology transfer from the CoABS "Grid" to and from the other grid-like projects. Cybenko had earlier met with NASA Ames to discuss coordination of the CoABS Grid with NASA's Information Power Grid. Visited with Dennis Gannon and Bill Johnston about possible relationships between the two efforts.

Human resources. One student completed a Ph.D thesis in the project, three undergraduates completed a Senior Honors thesis, and at least 11 other undergraduate students were involved in the research.

[5] http://www.gridforum.org

15 Kerf (David Kotz, Javed Aslam, Daniela Rus)

The objective of the Kerf project is to provide administrators with new methods for the analysis of an attack on their computer system. Numerous intrusion-detection tools exist; our focus is on intrusion analysis, specifically, tools that help administrators to examine large amounts of host and network log data. The Kerf tools fit into the unexplored territory between current approaches that search log data without providing much context and those that report summary statistics about records within the logs.

Given that an intrusion has been detected, the Kerf tools help the administrator answer basic questions about the attack: How, when, and where did the intruder get in? What did the intruder do here? Where did the intruder come from? Did the intruder attack remote machines using my system? Answering these questions allow the administrator to close security holes, determine damage, and collect evidence that may lead to the discovery and capture of the intruder.

Specifically, the aim of the Kerf project is to build semi-automated tools which will allow computer experts and system administrators to: (1) identify the characteristics of an attack given data from network sensors, (2 develop a hypothesis about the nature and origin of the attack, (3) assist the user by automatically refining and extrapolating that hypothesis, (4) share that hypothesis with security managers from other sites, (5) test that hypothesis at those other sites and coordinate the results of testing, and (6) archive the data necessary for use as evidence in later law enforcement actions.

15.1 Activities and Findings

The Kerf approach contributes five key components to the process of intrusion analysis. First, Kerfs logging facility securely records log entries away from the client hosts that may be attacked. Second, the logs are stored in an indexed database for quick and sophisticated retrieval. Third, we designed a query language, called SawQL, for intrusion analysis; it allows the sysadmin to express analysis hypotheses to the Kerf tools. Fourth, a graphical user interface includes visualization modules that can display the results in a compact, meaningful view. Finally, the hypothesis engine helps to automate the process of generating, refining, expanding, extrapolating, and generalizing hypotheses.

Secure logging. Most hackers who have successfully compromised a system proceed to remove traces of their intrusion from the systems logs. Thus, it is important to securely store host and network logging information offhost. There are many approaches and existing software for secure real-time transfer of log data from a collection of hosts to a secure log server. Kerf can take advantage of any such mechanism. For the purposes of our prototype we implemented a secure logging host that can receive, decode and store logging information from multiple sources.

Database. Many intrusions involve multiple hosts, and the evidence for many intrusions may be seen in multiple types of logs. To support fast retrieval of relevant records, Kerf's logging host stores incoming log records in a database, indexing on important fields (such as host, facility, any IP address mentioned in the record, and any user name mentioned in the record). This approach also serves to isolate the log collection mechanism from the analysis mechanism, and to limit the amount of parsing, indexing, and searching that must be done within our analysis tool. The current implementation uses MySQL.

Domain-specific query language. Given the database of log records, the analyst could use SQL queries to search for relevant records. SawQL (pronounced SAW-quill) is our extension to SQL designed specifically to express a sysadmin's hypothesis about an attack with maximum flexibility, abstracting the schema and join semantics of the underlying database. SawQL is oriented towards extracting *sequences* of logged event records correlated either temporally or on variables corresponding to common record fields such as hostnames, IP addresses, ports, and user names.

Data organization and presentation. The centerpiece of the Kerf toolset is the *Landing* application, which provides a graphical interface to the sysadmin. Landing allows the user to enter SawQL queries, displays the results of the queries, and allows the user to provide feedback to the hypothesis engine.

Hypothesis engine. Given a SawQL query from the sysadmin, Kerf extracts and displays the matching sequences. The GUI allows the sysadmin to mark each sequence "suspicious" or "innocuous" and to indicate the interesting elements of each suspicious sequence. Using any feedback provided (not all sequences need be marked), the engine uses algorithms drawn from the machine-learning community to suggest new queries that better fit the suspicious data, aiding with hypothesis refinement.

Findings

It is too early to tell whether the Kerf tools will make it easier for sysadmins to analyze attacks on their networks and hosts. At the time of this writing we are preparing a set of user studies at multiple universities. In these studies we will collect a large amount of data from live hosts and networks, and ask the administrators of those systems to use Kerf to explore the data and to analyze a problem or attack they encountered. By observing the actions they take, and interviewing them afterward, we hope to learn more about whether and how these tools are usable.

15.2 Products

Web site. `http://kerf.cs.dartmouth.edu`

Software. The Kerf project has developed a suite of software tools for intrusion analysis. Although the tools have not yet been released to the general public, they are presently being tested by system administrators at Dartmouth College and Northeastern University. The following software is in beta stage:

- Landing, the application for submitting correlation queries in the SawQL language and viewing the results.
- Treeview, the application for adaptive display of Landing result sets.
- PatternHelper, the command-line (readline, ncurses-based) application for bootstrapping a set of patterns for parsing free form syslog records.

Kerf's data analysis procedures and tools, especially adaptive data organization algorithms, could be incorporated in existing commercial analysis environments. We prepared a white paper for HSARPA addressing these possibilities. We have also begun a collaborative relationship with Core Security Technologies, makers of the Core Impact tool for vulnerability analysis, and there may eventually be opportunities for technology transfer.

Papers. The Kerf project resulted in one journal paper [ABK+04] and one workshop paper [ACKR01].

Publications

[ABK+04] Javed Aslam, Sergey Bratus, David Kotz, Ron Peterson, Daniela Rus, and Brett Tofel. The Kerf toolkit for intrusion analysis. *IEEE Security and Privacy*, 2(6):42–52, November/December 2004.

[ACKR01] Jay Aslam, Marco Cremonini, David Kotz, and Daniela Rus. Using mobile agents for analyzing intrusion in computer networks. In *Proceedings of the Workshop on Mobile Object Systems at ECOOP 2001*, July 2001.

15.3 Contributions

So far, the Kerf project has contributed an architecture for and new way of thinking about intrusion analysis, a field that has as yet seen little academic research. Our primary contribution is to demonstrate ways to integrate machine learning and hypothesis refinement into the process. Given the increasing frequency and complexity of Internet attacks, we believe that the Kerf tools could contribute substantially to the academic literature on the subject, to the corporate products that support sysadmins, and to the government's need for better cybersecurity.

Human resources. We have trained two postdoctoral fellows, one Ph.D student, and three summer students, in the course of this project thus far.

16 Armada (David Kotz)

High-performance computing increasingly occurs on "computational grids" composed of heterogeneous and geographically distributed systems of computers, networks, and storage devices that collectively act as a single "virtual" computer. A key challenge in this environment is to provide efficient access to data distributed across remote data servers. We explored some of the issues associated with I/O for wide-area distributed computing and developed an I/O system, called Armada, with the following features: a framework to allow application and dataset providers to flexibly compose graphs of processing modules that describe the distribution, application interfaces, and processing required of the dataset before or after computation; an algorithm to restructure application graphs to increase parallelism and to improve network performance in a wide-area network; and a hierarchical graph-partitioning scheme that deploys components of the application graph in a way that is both beneficial to the application and sensitive to the administrative policies of the different administrative domains. Our experiments showed that applications using Armada performed well in both low- and high-bandwidth environments, and that our approach does an exceptional job of hiding the network latency inherent in grid computing.

16.1 Activities and Findings

Unlike traditional parallel computers, grid applications execute in environments with unavoidable latency, low bandwidth, and unpredictable behavior. Our data-flow-based solution allows the application programmer and the dataset provider describe and deploy a network of application-specific and dataset-specific functionality across the grid. The application (or application library) controls virtually all aspects of the I/O system through a distributed graph of application components, including the application interface, optimization policies like caching and prefetching, and remote filtering of datasets. Armada supports remote execution by allowing the different components to execute on processors used by the client application, on processors used by storage servers, or on intermediate processors in the network. Our system automatically restructures the graph to distribute data flow and computation throughout the graph, and it places individual components using a scheme that is both beneficial to the application and considerate of administrative-domain allocation policies.

Our approach demonstrates that a flexible design along with careful attention to data-flow performance can lead to efficient I/O for grid applications. Our performance results show that the data-flow model does an exceptional job of hiding network latency inherent in grid computing. Applications using Armada perform well in low-bandwidth environments because restructured graphs allow an effective placement of Armada ships. Applications using Armada also perform well in high-bandwidth environments because restructured graphs often include end-to-end parallelism.

Outreach. We were involved with the Grid Forum from its founding, and Ph.D student Ron Oldfield was one of the key players in the data working group.

16.2 Products

Web site. The Armada project is described at http://www.cs.dartmouth.edu/~dfk/armada/. The parallel I/O archive, a community resource including an extensive bibliography, is also available at http://www.cs.dartmouth.edu/pario/.

Papers. The Armada project produced one conference paper [OK01a], two journal papers [OK02a, OK04], one invited book chapter [OK01b], one Ph.D thesis [Old03a], and three technical reports [OK02b, OK98, Old01]. Also, we wrote two encyclopedia chapters at the request of the editors; the first appeared [KJ99] and the second seems to be experiencing an indefinite publication delay [Kot02].

Publications

[KJ99] David Kotz and Ravi Jain. I/O in parallel and distributed systems. In Allen Kent and James G. Williams, editors, *Encyclopedia of Computer Science and Technology*, volume 40, pages 141–154. Marcel Dekker, Inc., 1999. Supplement 25.

[Kot02] David Kotz. Parallel input/output. In Joseph Urban and Partha Dasgupta, editors, *Encyclopedia of Distributed Computing*. Kluwer Academic Publishers, 2002. Accepted for publication.

[OK98] Ron Oldfield and David Kotz. Applications of parallel I/O. Technical Report PCS-TR98-337, Dept. of Computer Science, Dartmouth College, August 1998. Supplement to PCS-TR96-297.

[OK01a] Ron Oldfield and David Kotz. Armada: A parallel file system for computational grids. In *Proceedings of the First IEEE/ACM International Symposium on Cluster Computing and the Grid*, pages 194–201, Brisbane, Australia, May 2001. IEEE Computer Society Press.

[OK01b] Ron Oldfield and David Kotz. Scientific applications using parallel I/O. In Hai Jin, Toni Cortes, and Rajkumar Buyya, editors, *High Performance Mass Storage and Parallel I/O: Technologies and Applications*, chapter 45, pages 655–666. IEEE Computer Society Press and John Wiley & Sons, 2001.

[OK02a] Ron Oldfield and David Kotz. Armada: a parallel I/O framework for computational grids. *Future Generation Computing Systems (FGCS)*, 18(4):501–523, March 2002.

[OK02b] Ron Oldfield and David Kotz. Using the Emulab network testbed to evaluate the Armada I/O framework for computational grids. Technical Report TR2002-433, Dept. of Computer Science, Dartmouth College, Hanover, NH, September 2002.

[OK04] Ron Oldfield and David Kotz. Improving data access for computational grid applications. *Cluster Computing, The Journal of Networks, Software Tools and Applications*, 2004. Accepted for publication.

[Old01] Ron Oldfield. Summary of existing and developing data grids. White paper for the Remote Data Access group of the Global Grid Forum, March 2001.

[Old03a] Ron Oldfield. *Efficient I/O for Computational Grid Applications*. PhD thesis, Dept. of Computer Science, Dartmouth College, May 2003. Available as Dartmouth Computer Science Technical Report TR2003-459.

[Old03b] Ron Oldfield. Efficient I/O for computational grid applications. Technical Report TR2003-459, Dept. of Computer Science, Dartmouth College, May 2003.

16.3 Contributions

The primary contributions of this work are the following:

- a flexible framework, based on a data-flow programming model, that allows the application programmer and the dataset provider to deploy a network of application-specific and dataset-specific functionality across the grid;
- an algorithm to restructure a data-flow application graph to improve data flow across a wide-area network, based on programmer- and user-assigned properties that describe the behavior of the nodes within the graph;

- a hierarchical graph-partitioning scheme that leverages existing software to decide where to place individual application components in a way that benefits the application and abides by allocation policies set by individual administrative domains; and
- an evaluation of the I/O performance of a variety of applications using Armada.

The ideas developed in the Armada project should provide important insights in the emerging area of Grid computing, enabling computational scientists to be able to work with large remote data sets more efficiently. Their applications, ranging from weather modeling to oil exploration to nuclear weapons, have a broad effect on society.

Human resources. The Armada project represents the Ph.D dissertation work of Ron Oldfield, who is now on the research staff at Sandia National Laboratory.

17 Snowflake (David Kotz)

In the Snowflake project we tackled the problem of naming and sharing resources across administrative boundaries. Conventional systems manifest the hierarchy of typical administrative structure in the structure of their own mechanism. While natural for communication that follows hierarchical patterns, such systems interfere with naming and sharing that cross administrative boundaries, and therefore cause headaches for both users and administrators. Our work organized resource naming and security, not around administrative domains, but around the sharing patterns of users.

17.1 Activities and Findings

Our Snowflake project included four main parts.

First, we explored the challenges and tradeoffs involved in naming resources and considered a variety of existing approaches to naming.

Second, we considered the architectural requirements for user-centric sharing. We evaluated existing systems with respect to these requirements.

Third, to support the sharing architecture, we developed a formal logic of sharing that captures the notion of *restricted delegation*. Restricted delegation ensures that users can use the same mechanisms to share resources consistently, regardless of the origin of the resource, or with whom the user wishes to share the resource next. A formal semantics gives unambiguous meaning to the logic. We applied the formalism to the Simple Public Key Infrastructure (SPKI) and identified how the formalism either supports or discourages potential extensions to such a system.

Finally, we used the formalism to drive a user-centric sharing implementation for distributed systems. We showed how this implementation enables *end-to-end authorization*, a feature that makes heterogeneous distributed systems more secure and easier to audit. Conventionally, gateway services that bridge administrative domains, add abstraction, or translate protocols typically impede the flow of authorization information from client to server. In contrast, end-to-end authorization enables us to build gateway services that preserve authorization information; hence we reduce the size of the trusted computing base and enable more effective auditing.

17.2 Products

The Snowflake project produced two conference papers [HK00a, HK00b], four technical reports [HK98, HK99, HK00c, How98], a Ph.D dissertation [How00], one unrefereed newsletter article [HK00d], and one remaining unpublished paper [HK01].

Publications

[HK98] Jon Howell and David Kotz. Snowflake: Spanning administrative domains. Technical Report PCS-TR98-343, Dept. of Computer Science, Dartmouth College, December 1998.

[HK99] Jon Howell and David Kotz. An access-control calculus for spanning administrative domains. Technical Report PCS-TR99-361, Dept. of Computer Science, Dartmouth College, November 1999.

[HK00a] Jon Howell and David Kotz. End-to-end authorization. In *Proceedings of the 2000 Symposium on Operating Systems Design and Implementation*, pages 151–164. USENIX Association, October 2000.

[HK00b] Jon Howell and David Kotz. A formal semantics for SPKI. In *Proceedings of the Sixth European Symposium on Research in Computer Security (ESORICS 2000)*, volume 1895 of *Lecture Notes in Computer Science*, pages 140–158. Springer-Verlag, October 2000.

[HK00c] Jon Howell and David Kotz. A formal semantics for SPKI. Technical Report TR2000-363, Dept. of Computer Science, Dartmouth College, March 2000.

[HK00d] Jon Howell and David Kotz. Restricted delegation: seamlessly spanning administrative boundaries. *ACM Operating Systems Review*, 34(2):38–39, April 2000.

[HK01] Jon Howell and David Kotz. A formal semantics for SPKI. Unpublished., November 2001.

[How98] Jon Howell. Straightforward java persistence through checkpointing. Technical Report PCS-TR98-330, Dept. of Computer Science, Dartmouth College, May 1998.

[How00] Jonathan R. Howell. *Naming and sharing resources across administrative boundaries*. PhD thesis, Dept. of Computer Science, Dartmouth College, June 2000. Available as Dartmouth Computer Science Technical Reports TR2000-378, 379, and 380.

17.3 Contributions

We enumerated the naming and sharing qualities that relate to the goal of spanning administrative boundaries. We evaluate our architecture and existing architectures with respect to these qualities, and discovered that our approach trades off performance and user cost of storage management to enable names with high mnemonic and semantic value that users can easily share.

We contributed a naming mechanism based on user-relative paths that reflects user-to-user relationships. To ensure that all applications exhibit the benefits of user-specific naming, we structured the system so that naming is a separate, user-controlled layer between applications and other system services.

We contributed an authorization mechanism that enables users to uniformly specify their sharing requirements with other users, regardless of whether their colleagues are in the same administrative domain. Hence sharing reflects user-to-user relationships, not administrative hierarchy. Our sharing model was founded on a formal logic and semantics, so its meaning is unambiguous and implementations can be verified against a clear standard.

We established that our sharing model enables an end-to-end approach to authorization that has benefits even within administrative domains. It enabled us to build gateways that span network scales, levels of abstraction, and protocols while maintaining the flow of authorization information from the client to the ultimate resource server.

Our prototype system and applications demonstrated the naming and sharing mechanisms at work. We compared them to conventionally-organized systems and applications, and evaluated their characteristics qualitatively and quantitatively. The system exhibited the qualities we desire, and its performance roughly tracked that of conventional hop-by-hop authorization protocols with similar implementations.

The user-centered philosophy of system organization was the organizing element behind our work. We concluded that the philosophy is compatible with the usual goals of system design, and in fact simplifies the organization of systems by reducing many administrative tasks to special applications of user tools.

These results contributed to the scientific community interested in distributed systems, operating systems, and computer security. We hope that system architects will consider adopting our philosophy when they develop future designs.

Human resources. This project represents the Ph.D dissertation of Jon Howell, who is now on the research staff at Microsoft Research.

18 Solar (David Kotz)

Our research in mobile agents (see Section 14) led us to think about new ways to use mobile code to support information gathering, processing, and dissemination. The result was the "Solar" project, which aimed to support context-aware applications in pervasive computing.

The complexity of developing context-aware pervasive-computing applications calls for distributed software infrastructures that assist applications to collect, aggregate, and disseminate contextual data. We designed, implemented, and evaluated a Context Fusion Network (CFN), called Solar, which is built with a scalable and self-organized service overlay. Solar is flexible and allows applications to select distributed data sources and compose them with customized data-fusion operators into a directed acyclic information flow graph. Such a graph represents how an application computes high-level understandings of its execution context from low-level sensory data. To manage application-specified operators on a set of overlay nodes called Planets, Solar provides several unique services such as application-level multicast with policy-driven data reduction to handle buffer overflow, context-sensitive resource discovery to handle environment dynamics, and proactive monitoring and recovery to handle common failures.

Solar [Che04a] is a middleware infrastructure with two kinds of clients: *sensors* as data sources and *applications* as data sinks. A sensor may publish a data stream, by pushing data items called *events* into Solar. Some sensors also may have a pull-based interface, allowing users to query its current state. Applications ask Solar to find specified sensors and to execute application-supplied data-fusion *operators* to compute context. An operator is an independent data processing module that takes one or more data sources as input and acts as another data source.

18.1 Activities and Findings

We implemented two Solar prototypes, both in Java. Both prototypes adopted an operator composition programming model and similar design choices [CK02a]. We implemented the first prototype, with a centralized architecture for simplicity, for two "pervasive-computing" seminar courses in which Solar was used by students to develop applications [CK01a, CK02d]. Our experience with the first prototype, including an analysis of a sensor environment [CK04a], performance and interoperability [Whi02a, Whi02b], the security and access control design [MK02, MK05a, MK05b, Mas02], and several application studies [Mat01, WCK04], contributed to the design and implementation of a second version of Solar [CLK04]. The second prototype used a fully distributed and self-organized architecture, and the software package consisted of more than 13,000 lines of code. The later version also makes several research contributions on its generalizable operator-management services [CK03, CK05, CK04d].

Findings

Solar provides a suite of services to manage these operators and to meet the challenges of a heterogeneous and dynamic pervasive-computing environment. The primary contributions of Solar are:

- a flexible and scalable context-fusion system for a pervasive computing environment with a programming model based on an operator graph compositional model. This model allows applications to deploy specific data-fusion functionality inside the network;
- a policy-driven data-reduction technique that allows loss-tolerant context-aware applications to trade completeness for fast delivery in case of buffer overflows caused by rapid data streams and slow receivers;
- a scalable naming service that supports persistent queries and context-sensitive resource discovery, which takes advantage of our context-fusion facility and allows applications to off-load most traffic and computation into the infrastructure;

- a set of component monitoring and dependency tracking protocols that automatically recover operators that are lost due to host failures and that automatically adjust the operator graphs according to changes in resource availability;
- a detailed analysis of several traces collected from location tracking systems that are in daily usage, which provides insights on typical pervasive-computing characteristics; and
- experience building pervasive-computing applications and lessons learned. Together with an open-source software package, Solar provides valuable contributions to the community.

One barrier preventing pervasive-computing from faster adoption is the lack of a flexible and scalable infrastructure designed to meet the challenges of a heterogeneous and volatile environment. To support context-aware applications, our Solar system employs a flexible operator-composition programming model that facilitates both development and deployment of these applications. To manage application-specified operators on a set of overlay nodes called Planets, Solar provides several unique services such as application-level multicast with policy-driven data reduction to handle buffer overflow, context-sensitive resource discovery to handle environment dynamics, and proactive monitoring and recovery to handle common failures. Experimental results show that these services perform well on a typical DHT-based peer-to-peer routing substrate. Our experience from building applications with Solar shows that developers benefit from our infrastructure, allowing them to focus on task-specific functionalities without being overwhelmed by the complexity of a pervasive-computing environment.

18.2 Products

Web site. http://www.cs.dartmouth.edu/~solar/

Software. We distribute the source code for the Solar system on our web site.

Papers. The Solar project resulted in one journal paper [MK05a], eight refereed conference and workshop papers [CLK04, CK05, CK03, CK02a, CK02d, CK01a, MK05b, WCK04], eight technical reports [CK04a, CK00, CK02b, CK04c, CK04b, CK02c, CK01b, MK02, Whi02b], three undergraduate senior theses [Mas02, Mat01, Whi02a], one Ph.D thesis so far [Che04a], one M.S. thesis [Wan04], and one poster abstract [CK04d].

Publications

[Che04a] Guanling Chen. *Solar: Building A Context Fusion Network for Pervasive Computing*. PhD thesis, Dept. of Computer Science, Dartmouth College, August 2004.

[Che04b] Guanling Chen. Solar: Building a context fusion network for pervasive computing. Technical Report TR2004-514, Dept. of Computer Science, Dartmouth College, August 2004.

[CK00] Guanling Chen and David Kotz. A survey of context-aware mobile computing research. Technical Report TR2000-381, Dept. of Computer Science, Dartmouth College, November 2000.

[CK01a] Guanling Chen and David Kotz. Solar: Towards a flexible and scalable data-fusion infrastructure for ubiquitous computing. In *UbiTools workshop at UbiComp 2001*, October 2001.

[CK01b] Guanling Chen and David Kotz. Supporting adaptive ubiquitous applications with the SOLAR system. Technical Report TR2001-397, Dept. of Computer Science, Dartmouth College, May 2001.

[CK02a] Guanling Chen and David Kotz. Context aggregation and dissemination in ubiquitous computing systems. In *Proceedings of the Fourth IEEE Workshop on Mobile Computing Systems and Applications*, pages 105–114. IEEE Computer Society Press, June 2002.

[CK02b] Guanling Chen and David Kotz. Context aggregation and dissemination in ubiquitous computing systems. Technical Report TR2002-420, Dept. of Computer Science, Dartmouth College, February 2002.

[CK02c] Guanling Chen and David Kotz. Solar: A pervasive-computing infrastructure for context-aware mobile applications. Technical Report TR2002-421, Dept. of Computer Science, Dartmouth College, February 2002.

[CK02d] Guanling Chen and David Kotz. Solar: An open platform for context-aware mobile applications. In *Proceedings of the First International Conference on Pervasive Computing (Short paper)*, pages 41–47, June 2002. In an informal companion volume of short papers.

[CK03] Guanling Chen and David Kotz. Context-aware resource discovery. In *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications*, pages 243–252. IEEE Computer Society Press, March 2003.

[CK04a] Guanling Chen and David Kotz. A Case Study of Four Location Traces. Technical Report TR2004-490, Dept. of Computer Science, Dartmouth College, February 2004.

[CK04b] Guanling Chen and David Kotz. Application-controlled loss-tolerant data dissemination. Technical Report TR2004-488, Dept. of Computer Science, Dartmouth College, February 2004.

[CK04c] Guanling Chen and David Kotz. Dependency management in distributed settings. Technical Report TR2004-495, Dept. of Computer Science, Dartmouth College, March 2004.

[CK04d] Guanling Chen and David Kotz. Dependency management in distributed settings (poster abstract). In *International Conference on Autonomic Computing (ICAC-04)*, May 2004.

[CK05] Guanling Chen and David Kotz. Policy-driven data dissemination for context-aware applications. In *Proceedings of the Third IEEE International Conference on Pervasive Computing and Communications*, March 2005. Short paper.

[CLK04] Guanling Chen, Ming Li, and David Kotz. Design and implementation of a large-scale context fusion network. In *First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services (Mobiquitous)*, pages 246–255, August 2004.

[Mas02] Chris Masone. Role definition language (RDL): A language to describe context-aware roles. Technical Report TR2002-426, Dept. of Computer Science, Dartmouth College, May 2002.

[Mat01] Arun Mathias. SmartReminder: A case study on context-sensitive applications. Technical Report TR2001-392, Dept. of Computer Science, Dartmouth College, June 2001. Senior Honors Thesis.

[MK02] Kazuhiro Minami and David Kotz. Controlling access to pervasive information in the "Solar" system. Technical Report TR2002-422, Dept. of Computer Science, Dartmouth College, February 2002.

[MK05a] Kazuhiro Minami and David Kotz. Secure context-sensitive authorization. *Journal of Pervasive and Mobile Computing*, 1(1), January 2005. Accepted for publication.

[MK05b] Kazuhiro Minami and David Kotz. Secure context-sensitive authorization. In *Proceedings of the Third IEEE International Conference on Pervasive Computing and Communications*, March 2005. Accepted for publication.

[Wan04] Jue Wang. Performance evaluation of a resource discovery service. Technical Report TR2004-513, Dept. of Computer Science, Dartmouth College, October 2004.

[WCK04] Jue Wang, Guanling Chen, and David Kotz. A sensor fusion approach for meeting detection. In *MobiSys 2004 Workshop on Context Awareness*, June 2004.

[Whi02a] A. Abram White. Performance and interoperability in Solar. Technical Report TR2002-427, Dept. of Computer Science, Dartmouth College, May 2002.

[Whi02b] A. Abram White. XSLT and XQuery as operator languages. Technical Report TR2002-429, Dept. of Computer Science, Dartmouth College, May 2002.

18.3 Contributions

The Solar project has contributed a new approach for the distributed collection, processing, and dissemination of sensor information for context-aware pervasive-computing applications. We list specific contributions in the Findings section above.

We are now adapting and applying the Solar middleware to a new application domain: situational awareness for emergency responders in large-scale disaster areas. In this project funded by the Department of Homeland Security through Dartmouth's Institute for Security Technology Studies, we are working with two other groups. One group, led by Daniela Rus, is developing sensor-network technology for environmental monitoring in a disaster area; these sensor networks provide critical input to the Solar middleware. Another group, led by Sue McGrath, is developing physiological sensors for monitoring the health of both responders and victims (providing more data into Solar); her group also develops the command-and-control interfaces and applications as clients of the Solar system. We believe that this transfer of Solar technology to emergency response applications will be directly beneficial to society, as well as leading to new scientific insight.

Human resources. One Ph.D, one M.S., and three undergraduate theses resulted; another two Ph.D theses are expected.

19 Wireless networks (David Kotz)

We have an active group of students and postdocs interested in wireless networks, and we have focused on measurements of live production networks, experiments with real implementations of research protocols, and methods for comparing simulation and experimental results. All of this work, even the papers published since the end of the award period, benefited tremendously in the final six months of the award from new servers and storage that allow us to store and analyze a terabyte of network trace data.

19.1 Activities and Findings

We have been active in several areas. Again, the common theme is measurement, and experiments, with real networks and implementations.

Characterizing usage of the Dartmouth wireless LAN. Understanding usage patterns in wireless local-area networks (WLANs) is critical for those who develop, deploy, and manage WLAN technology, as well as those who develop systems and application software for wireless networks. When Dartmouth College installed an 802.11b ("WiFi") network in the spring of 2001, we decided to measure actual usage on the network. An undergraduate thesis student developed the initial data-collection infrastructure and collected a small round of data in the first months of the network's operation [Ste01]. Later, another student used a PDA to roam around campus to develop methods for mapping wireless signal strength [Len03].

Over the summer of 2001, another undergraduate student improved the data collection infrastructure and in the fall we conducted the largest and most comprehensive trace of network activity in a large, production wireless LAN. For eleven weeks we traced the activity of nearly two thousand users drawn from a general campus population, using a campus-wide network of 476 access points spread over 161 buildings at Dartmouth College. Our study expands on all earlier studies, with a significantly larger and broader population.

We published our results in Mobicom'02 [KE02a], followed by a technical report that corrected and expanded on that paper [KE02b], and a journal paper that significantly expanded the analysis and compared the Fall 2001 data with additional data collected in Spring 2002 [KE05].

We found that residential traffic dominated all other traffic, particularly in residences populated by newer students; students are increasingly choosing a wireless laptop as their primary computer. Although web protocols were the single largest component of traffic volume, network backup and file sharing contributed an unexpectedly large amount to the traffic. Although there was some roaming within a network session, we were surprised by the number of situations in which cards roamed excessively, unable to settle on one access point. Cross-subnet roams were an especial problem, because they broke IP connections, indicating the need for solutions that avoid or accommodate such roams.

In Fall 2003 we set out to repeat our study, with several goals in mind. First, we wanted to collect wireless packet traces from a wider area of campus, and to see whether and how usage had changed. Second, we wanted to focus on Voice-over-IP traffic on the wireless LAN (sometimes called Voice over WLAN or VoWLAN). Third, we wanted a better understanding of user mobility. This study included more than 550 access points and 7000 users over seventeen weeks. We employed several measurement techniques, including syslogs, telephone records, SNMP polling and tcpdump packet sniffing. This is still the largest WLAN study to date, and the first to look at a large, mature WLAN and consider geographic mobility. The result was a paper in Mobicom'04 [HKA04a], which has been extended in a technical report [HKA04b]; a journal paper is in preparation.

We found that the applications used on the WLAN changed dramatically between Fall 2001 and Fall 2003. Initial WLAN usage was dominated by Web traffic; the new trace shows significant increases in peer-to-peer, streaming multimedia, and voice over IP (VoIP) traffic. On-campus traffic now exceeds off-campus traffic, a reversal of the situation at the WLAN's initial deployment. Our study indicates that VoIP had been used little on the wireless network thus far, and most VoIP calls are made on the wired network. Most calls last less than a minute.

We saw more heterogeneity in the types of clients used, with more embedded wireless devices such as PDAs and mobile VoIP clients. We defined a new metric for mobility, the "session diameter." We used this metric to show that embedded devices have different mobility characteristics than laptops, and travel further and roam to more access points. Overall, users were surprisingly non-mobile, with half remaining close to home about 98% of the time.

Even with all of this data and careful analysis, we were often unable to truly understand *why* people used the network the way they did, or how that network activity correlated with other off-network activity. So, in the summer of 2004 we teamed with a sociologist to survey and monitor the students using the Experience Sampling Method, while simultaneously monitoring their traffic with our existing measurement infrastructure. We monitored 29 users remotely for one week, and signaled them to fill out a questionnaire whenever interesting wireless behavior was observed. In our paper, to appear soon, we found ESM to be a useful method for collecting data about wireless network usage that cannot be provided by network monitoring, and we presented a list of recommendations for network researchers who wish to conduct an ESM study [HAK05].

Voice over IP over wireless LAN. Voice is an increasingly important application for wireless networks, especially as WiFi handsets (or WiFi-cellular handsets) become available. In addition to our work to characterize the prevalence and behavior of voice traffic on the Dartmouth network, mentioned above [HKA04a, HKA04b], a few years earlier we experimented with methods for voice-call roaming in a wireless LAN. In her senior thesis, Ayorkor Mills-Tettey developed a fast call-handoff method for devices that roam across subnet boundaries, implemented it using the H.323 protocol stack, and conducted experiments to measure handoff latency [MT01]. She later published her thesis as a conference paper [MTK02], where it won the Best Student Paper award.

Simultaneously, we recognized the need for a distributed directory service to allow callers to find the desired party, even when the remote party was roaming across subnet boundaries; Ammar Khalid designed and implemented such a service (and integrated it with Ayorkor's system) in his senior honors thesis [Kha01].

Wireless network authentication. In the new standards for WLAN security, many choices exist for the authentication process. We wrote a survey paper that lists eight desired properties of WLAN authentication protocols, surveys eight recent authentication protocols, and analyzes the protocols according to the desired properties [BSK04].

Mobile ad hoc networks. Most comparisons of wireless ad hoc routing algorithms involve simulated or *indoor* trial runs, or outdoor runs with only a small number of nodes, potentially leading to an incorrect picture of algorithm performance. In our work, we conducted an outdoor comparison of four different routing algorithms, APRL, AODV, ODMRP, and STARA, running on top of thirty-three 802.11-enabled laptops moving randomly through an athletic field. This comparison provides insight into the behavior of ad hoc routing algorithms at larger real-world scales than have been considered so far. In addition, we compared the outdoor results with both indoor ("tabletop") and simulation results for the same algorithms, examining the differences between the indoor results and the outdoor reality. We presented these results in a conference paper [GKN+04a] and more extensively in a technical report [GKN+04b].

In addition, we developed a software infrastructure that allowed us to implement the ad hoc routing algorithms and use the *same* codebase for indoor, outdoor, and simulated trial runs. This approach allowed us to directly validate wireless models against the outdoor experiments described above. The simulator read traces collected from the outdoor experiments, and used them to drive direct-execution implementations of the routing protocols. Because we were able to reproduce the same network conditions as in the real experiment, comparing the routing behavior *measured* in the real experiment with behavior *computed* by the simulation, we were able to validate the models of radio behavior upon which protocol behavior depends. We concluded that, contrary to popular belief, it is *possible* to have fairly accurate results using a simple wireless model (though not as simple as those often used in the literature). We observed, however, that the routing behavior is quite sensitive to one

of this model's parameters. The implication is that one should i) use a more complex wireless model that explicitly models point-to-point path loss, ii) use measurements from an environment typical of the one of interest, or iii) study behavior over a range of environments to identify the sensitivities of the protocol's performance under different network conditions. These results were published in a conference paper [LYN+04] and a journal paper [LYN+05].

From our experience implementing, measuring, and simulating ad hoc routing protocols, and from canvassing the literature in the field, we found that many of the assumptions commonly used in most prior research are simply not valid, and that the (often implicit) use of these axioms can lead to questionable results. In a conference paper [KNG+04] and extended technical report [KNE03] we provide a comprehensive review of six assumptions that are still part of many ad hoc network simulation studies, despite increasing awareness of the need to represent more realistic features, including hills, obstacles, link asymmetries, and unpredictable fading. We used our measurements from the above routing experiments to demonstrate the weakness of these assumptions, and showed how these assumptions can cause simulation results to differ significantly from experimental results. We developed a series of recommendations for researchers, whether they develop protocols, analytic models, or simulators for ad hoc wireless networks.

A key player in all of these studies was undergraduate Cal Newport, who spent much of his junior and senior year helping to conduct the experiments and analyze the data. His senior honors thesis [New04] combines many of the results from the papers above. He is now a Ph.D student in computer science at MIT, and is preparing a journal submission based on his thesis.

Mobility modeling and prediction. There is increasing interest in location-aware systems and applications. It is important for any designer of such systems and applications to understand the nature of user and device mobility. Furthermore, an understanding of the effect of user mobility on access points (APs) is also important for designing, deploying, and managing wireless networks. Using our data collected about several thousand Dartmouth wireless network users over three years, we are able to study mobility patterns in far greater detail than most prior studies.

Our initial work, undergraduate Clara Lee's senior honors thesis [Lee03], attempted to apply the metrics of persistence and prevalence (developed by Balazinska at IBM). We later learned of flaws in the measurement data [HK03], however, that invalidate these results. Thus, we collected fresh data and examine it more closely in our Mobicom 04 paper [HKA04a], and an extended technical report [HKA04b]. In that paper we also introduce a new metric, the "session diameter," that allows us to compare the degree of mobility across users, locations, or days.

More recently, we developed a general methodology for extracting mobility information from wireless network traces, and for classifying mobile users and APs [KK05]. We used the Fourier transform to convert time-dependent location information to the frequency domain, then chose the two strongest periods and used them as parameters to a classification system based on Bayesian theory. We found that user mobility had a strong period of one day, but there was also a large group of users that had either a much smaller or much bigger primary period. Both primary and secondary periods had important roles in determining classes of mobile users. Users with one day as their primary period and a smaller secondary period were most prevalent; we expect that they were mostly students taking regular classes. Because most APs had one day as their strongest period, the secondary period played a critical role in classifying the APs. APs with one day as their primary period and one week as their secondary period were most prevalent. By plotting the classes of APs on our campus map, we discovered that this periodic behavior of APs seemed to be independent of their geographical locations, but may depend on the relative locations of nearby APs.

Finally, we examined the potential for user mobility to be predictable, since there is some indication that some location-aware applications and wireless networks can better serve their clients by anticipating client mobility. Many location predictors have been proposed in the literature, though few have been evaluated with empirical evidence. Our work [SKJH04a, SKJH04b] reports the results of the first extensive empirical evaluation of location predictors. We implemented and compared the prediction accuracy of several location predictors

drawn from four major families of domain-independent predictors, namely Markov-based, compression-based, PPM, and SPM predictors. We found that low-order Markov predictors performed as well or better than the more complex and more space-consuming compression-based predictors. Predictors of both families fail to make a prediction when the recent context has not been previously seen. To overcome this drawback, we added a simple fallback feature to each predictor and found that it significantly enhanced its accuracy in exchange for modest effort. Thus the Order-2 Markov predictor with fallback was the best predictor we studied, obtaining a median accuracy of about 72% for users with long trace lengths. We also investigated a simplification of the Markov predictors, where the prediction is based not on the most frequently seen context in the past, but the most recent, resulting in significant space and computational savings. We found that Markov predictors with this recency semantics can rival the accuracy of standard Markov predictors in some cases. Finally, we considered several seemingly obvious enhancements, such as smarter tie-breaking and aging of context information, and discovered that they had little effect on accuracy.

19.2 Products

Web site and data collection. We created a simple website to explain our wireless characterization studies.[6] We cleaned and anonymized much of our data, and make it available on the web to any academic research, and to our corporate research partners; as a service to the community we also host mirrors of other prominent datasets from other wireless-measurement efforts.[7]

Papers. The work described in this section produced nine refereed conference and workshop papers [GKN+04a, HAK05, HKA04a, KK05, KE02a, KNG+04, LYN+04, MTK02, SKJH04a], two refereed journal papers [KE05, LYN+05], seven technical reports [BSK04, GKN+04b, HK03, HKA04b, KE02b, KNE03, SKJH04b], and six undergraduate honors theses [Lee03, Len03, Kha01, MT01, New04, Ste01]. Four journal papers are in preparation.

Publications

[BSK04] Kwang-Hyun Baek, Sean W. Smith, and David Kotz. A survey of WPA and 802.11i RSN authentication protocols. Technical Report TR2004-524, Dept. of Computer Science, Dartmouth College, Hanover, NH, November 2004.

[GKN+04a] Robert S. Gray, David Kotz, Calvin Newport, Nikita Dubrovsky, Aaron Fiske, Jason Liu, Christopher Masone, Susan McGrath, and Yougu Yuan. Outdoor experimental comparison of four ad hoc routing algorithms. In *Proceedings of the ACM/IEEE International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM)*, pages 220–229, October 2004. Finalist for Best Paper award.

[GKN+04b] Robert S. Gray, David Kotz, Calvin Newport, Nikita Dubrovsky, Aaron Fiske, Jason Liu, Christopher Masone, Susan McGrath, and Yougu Yuan. Outdoor experimental comparison of four ad hoc routing algorithms. Technical Report TR2004-511, Dept. of Computer Science, Dartmouth College, 2004.

[HAK05] Tristan Henderson, Denise Anthony, and David Kotz. Measuring wireless network usage with the experience sampling method. In *Proceedings of the First Workshop on Wireless Network Measurements (WiNMee)*, April 2005.

[6] http://www.cs.dartmouth.edu/~campus/.
[7] http://cmc.cs.dartmouth.edu/data/.

[HK03] Tristan Henderson and David Kotz. Problems with the Dartmouth wireless SNMP data collection. Technical Report TR2003-480, Dept. of Computer Science, Dartmouth College, December 2003.

[HKA04a] Tristan Henderson, David Kotz, and Ilya Abyzov. The changing usage of a mature campus-wide wireless network. In *Proceedings of the Tenth Annual International Conference on Mobile Computing and Networking*, pages 187–201. ACM Press, September 2004.

[HKA04b] Tristan Henderson, David Kotz, and Ilya Abyzov. The changing usage of a mature campus-wide wireless network. Technical Report TR2004-496, Dept. of Computer Science, Dartmouth College, March 2004.

[KE02a] David Kotz and Kobby Essien. Analysis of a campus-wide wireless network. In *Proceedings of the Eighth Annual International Conference on Mobile Computing and Networking*, pages 107–118, September 2002. Revised and corrected as Dartmouth CS Technical Report TR2002-432.

[KE02b] David Kotz and Kobby Essien. Analysis of a campus-wide wireless network. Technical Report TR2002-432, Dept. of Computer Science, Dartmouth College, September 2002.

[KE05] David Kotz and Kobby Essien. Analysis of a campus-wide wireless network. *Wireless Networks*, 11:115–133, 2005.

[Kha01] Ammar Khalid. A directory infrastructure to support mobile services. Technical Report TR2001-391, Dept. of Computer Science, Dartmouth College, June 2001. Senior Honors Thesis.

[KK05] Minkyong Kim and David Kotz. Classifying the mobility of users and the popularity of access points. In *Proceedings of the International Workshop on Location- and Context-Awareness (LoCA)*, Lecture Notes in Computer Science. Springer-Verlag, May 2005.

[KNE03] David Kotz, Calvin Newport, and Chip Elliott. The mistaken axioms of wireless-network research. Technical Report TR2003-467, Dept. of Computer Science, Dartmouth College, July 2003.

[KNG+04] David Kotz, Calvin Newport, Robert S. Gray, Jason Liu, Yougu Yuan, and Chip Elliott. Experimental evaluation of wireless simulation assumptions. In *Proceedings of the ACM/IEEE International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM)*, pages 78–82, October 2004.

[Lee03] Clara Lee. Persistence and prevalence in the mobility of dartmouth wireless network users. Technical Report TR2003-455, Dept. of Computer Science, Dartmouth College, May 2003.

[Len03] Chris Lentz. 802.11b wireless network visualization and radiowave propagation modeling. Technical Report TR2003-451, Dept. of Computer Science, Dartmouth College, June 2003.

[LYN+04] Jason Liu, Yougu Yuan, David M. Nicol, Robert S. Gray, Calvin C. Newport, David Kotz, and Luiz Felipe Perrone. Simulation validation using direct execution of wireless ad-hoc routing protocols. In *Proceedings of the Workshop on Parallel and Distributed Simulation (PADS)*, pages 7–16, May 2004. Nominated for Best Paper award.

[LYN+05] Jason Liu, Yougu Yuan, David M. Nicol, Robert S. Gray, Calvin C. Newport, David Kotz, and Luiz Felipe Perrone. Validation of wireless models using direct-execution simulation of ad-hoc routing protocols. *SIMULATION: Transactions of The Society for Modeling and Simulation International*, January 2005. Accepted for publication in the "Best of PADS 2004" special issue.

[MT01] G. Ayorkor Mills-Tettey. Mobile voice over IP (MVOIP): An application-level protocol. Technical Report TR2001-390, Dept. of Computer Science, Dartmouth College, June 2001. Senior Honors Thesis.

[MTK02] G. Ayorkor Mills-Tettey and David Kotz. Mobile voice over IP (MVOIP): An application-level protocol for call hand-off in real time applications. In *Proceedings of the Twenty-first IEEE International Performance, Computing, and Communications Conference*, pages 271–279. IEEE Computer Society Press, April 2002.

[New04] Calvin Newport. Simulating mobile ad hoc networks: a quantitative evaluation of common MANET simulation models. Technical Report TR2004-504, Dept. of Computer Science, Dartmouth College, June 2004. Senior Honors Thesis.

[SKJH04a] Libo Song, David Kotz, Ravi Jain, and Xiaoning He. Evaluating location predictors with extensive Wi-Fi mobility data. In *Proceedings of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, March 2004.

[SKJH04b] Libo Song, David Kotz, Ravi Jain, and Xiaoning He. Evaluating location predictors with extensive Wi-Fi mobility data. Technical Report TR2004-491, Dept. of Computer Science, Dartmouth College, February 2004.

[Ste01] Pablo Stern. Measuring early usage of Dartmouth's wireless network. Technical Report TR2001-393, Dept. of Computer Science, Dartmouth College, June 2001. Senior Honors Thesis.

19.3 Contributions

Our wireless usage characterization includes the two largest-ever studies of a wireless LAN in operation, examining the activity of thousands of users over several months (or years, in some papers) on over 500 access points on a complete campus-wide network. The methodology (and software) we developed for our work has been re-used elsewhere to gather and analyze data in other settings, including other campuses and corporate offices. The data we developed for our work is in use at 44 other university and corporate research labs. The result is that the wireless research community is much more deeply grounded with an understanding of the way that real users and real networks behave. Furthermore, the data and analysis should allow network providers and vendors to improve the quality of the products used by wireless computer users everywhere.

Similarly, our work on mobility prediction is unusual in its use of real data, so we were able to quantitatively compare a wide range of prediction algorithms for the first time. This work has applications in wireless network management, particularly for VoWLAN, and should allow network providers to improve service quality for voice customers.

Our survey paper on wireless network authentication protocols is the only comprehensive study and comparison of these protocols; we believe the community benefits from a careful description and cross-comparison as they develop new protocols or attempt to understand the value of each protocol.

Finally, our work in ad hoc networks includes the largest-ever outdoor experiment using real implementations of ad hoc routing algorithms and a real wireless network. We were thus able to quantitatively compare routing algorithms that are usually compared only in simulations, *and* we were able to compare our experimental results to those found in simulation. The results include a broad series of recommendations to the community, which desperately needs better simulators, better models, and more careful experimental methods.

Human resources. This work has, so far, led to six undergraduate Senior Honors Theses, and at least two Ph.D theses will come of it. Two other undergraduate students were significantly involved, and are co-authors of the Mobicom papers. Indeed, most of the network-characterization research has been conducted by undergraduates, with the recent addition of a postdoctoral research fellow (Tristan Henderson). The mobility prediction work involves two graduate students. The work with ad hoc networks involves collaborators from other institutions, and at the time included two graduate students, several undergraduate students, and several research staff at Dartmouth, University of Illinois, and BBN Labs.

20 Biomedical Computing (Fillia Makedon)

Biomedical computing research in the Dartmouth Experimental Visualization Laboratory (DEVLAB), Fillia Makedon's lab, has focused on several areas: neuroscience, cardiology, and genetics. Specific projects have focused on

1. computational methods for finding functional and structural determinants of neurobehavioral disorders,
2. building and sharing access to secure databases for patient information,
3. analyzing time series images of the heart and lungs to correct for motion artifacts and model systematic behaviors, and
4. computational methods for determining and evaluating marker genes for various conditions or diseases.

20.1 Activities and Findings

1. **Efficient classification methods for individual brain activation maps**
   - Development of novel classification techniques for three-dimensional region data.
   - Development of efficient and effective techniques for characterizing image data and facilitating similarity queries or content-based retrieval.
   - Initial attempts to construct of a framework for evaluating many aspects of association mining methods in medical image databases, and in particular methods that can be used to discover associations between brain activations and tasks performed.
   - Investigation of Fisher Linear Discriminant (FLD) analysis as a non-threshold based tool for comparing brain activation maps.
   - Integration of structural information, including structure volume as well as structure shape and lesion data. We are currently investigating the extension of our data mining approach to include structural brain information.
2. **Efficient three-dimensional shape analysis using parametric surface modeling for regions of interest in structural brain maps**
   - Development of a new framework for 3D surface object classification that combines a powerful shape description method with a set of effective pattern classification, feature selection, evaluation, and visualization techniques;
   - Development of a new multi-surface alignment algorithm that extends the above framework from the single object case to the multiple object case;
   - Development of a new spherical parameterizations algorithm that makes the above framework applicable to not only voxel surfaces but also general triangle meshes.
3. **Ongoing development of a secure Heart Failure Database**
   - Development of an extension to the secure clinical Computer Information System (CIS) at the Dartmouth-Hitchcock Medical Center in order to provide a more convenient data entry and retrieval platform
4. **Development of new registration and 3D motion modeling tools**
   - Creation of tools to cancel lung motion in images of pulmonary nodules

- Creation of spherical harmonic-based models of heart motion and application to simulating and predicting heart dynamics

5. **Biweekly educational seminars on biomedical topics**

20.2 Products

A partial list of products includes

1. The TROI system (Tracing Regions Of Interest): a tool a system for training specialists in the tracing of regions of interest in MRI images
2. The IVM system: an Interactive Vessel Manipulation tool that can help make effective and efficient assessment of angiogenesis and arteriogenesis in computed tomographic angiography (CTA) studies.
3. CAT Tree (Computer Aided Traversal Tree): a human-computer interface component to accelerate navigation of trees and tree-like data structures
4. HykGene: a clustering-based approach for isolating discriminative genes "marker genes") in microarray data
5. WSVD (Weighted Sum of Vector Distances): a feature extraction method for quantifying evolving processes in multimodal 3D medical images
6. a prototype searchable MRI database in which scan data is retrieved on the basis of the similarity of metadata, including image features like localized brightness or curvature

Publications

[ATK+03] L. G. Astrakas, A. A. Tzika, S. Kapadakis, F. Makedon, Song Ye, and J. Ford. The clinical perspective of large scale projects: A case study with pediatric brain tumors & multiparametric mr imaging. In *Human Computer Interaction*, Crete, Greece, 2003.

[CSM+01a] Ling Cheng, Li Shen, Fillia S. Makedon, Annette M. Donnelly, Laura A. Flashman, and Andrew J. Saykin. On the development of a training and research system for tracing human brain structures. In *SCI 2001: The 5th World Multi-Conference on Systemics, Cybernetics and Informatics*, volume VIII, pages 151–155, Orlando, Florida, USA, July 22–25, 2001.

[CSM+01b] Ling Cheng, Li Shen, Fillia S. Makedon, Annette M. Donnelly, Laura A. Flashman, and Andrew J. Saykin. TROI – a training and research system for tracing regions of interest in brain images. *NeuroImage*, 13(6):S1298, June 2001.

[FFM+03] J. Ford, H. Farid, F. Makedon, L. A. Flashman, T W McAllister, V. Megalooikonomou, and A. J. Saykin. Patient classification from fmri brain activation patterns. In *Sixth International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2003)*, Montreal, Canada, 2003.

[FMM+01] James Ford, Fillia Makedon, Vasileios Megalooikonomou, Li Shen, Tilmann Steinberg, and Andrew J. Saykin. Spatial comparison of fMRI activation maps for data mining. *NeuroImage*, 13(6):S1302, June 2001.

[For03] James Ford. *Patient Classification from fMRI Brain Activation Patterns*. PhD thesis, Dartmouth College, Hanover, NH, July 2003.

[FSM+02] James Ford, Li Shen, Fillia Makedon, Laura Flashman, and Andrew Saykin. A combined structural-functional classification of schizophrenia using hippocampal volume plus fMRI activation. In *EMBS-BMES2002 Second Joint Meeting of the IEEE Engineering in Medicine and Biology Society and the Biomedical Engineering Society*, volume 1, pages 48–49, Houston, Texas, October 23–26, 2002.

[HMP+04] Heng Huang, Fillia Makedon, Justin Pearlman, James Ford, Li Shen, Yuhang Wang, and Ling Gao. Efficient similarity retrieval framework for temporal shape sequences: A case study in cardiac MR images. In *Proc. 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Francisco, California, September 1, 2004.

[HSF+05] Heng Huang, Li Shen, James Ford, Fillia Makedon, Ling Gao, and Justin Pearlman. Functional analysis of cardiac MR images using spharm modeling. In *SPIE Medical Imaging 2005: Image Processing*, San Diego, California, February 12–17, 2005.

[HSM+04] Heng Huang, Li Shen, Fillia S. Makedon, Ling Gao, and Justin Pearlman. Three-dimensional analysis of cardiac magnetic resonance imaging using spherical harmonics model. In *ACC'04: Annual Scientific Session of the American College of Cardiology*, New Orleans, March 7–10, 2004.

[HSM+05] Heng Huang, Li Shen, Fillia Makedon, Sheng Zhang, Mark Greenberg, Ling Gao, and Justin Pearlman. A clustering-based approach for prediction of cardiac resynchronization therapy. In *20th ACM Symposium on Applied Computing*, Santa Fe, New Mexico, March 13–17, 2005.

[Mak00] F. Makedon. Tools towards mining human brain data (invited talk). In *World Conference on the WWW and Internet*, San Antonio, TX, 2000.

[MFSM00] Vasileios Megalooikonomou, James Ford, Li Shen, and Fillia Makedon. Data mining in brain imaging. *Statistical Methods in Medical Research*, 9(4):359–394, 2000.

[MTA+03] Fillia Makedon, A. Aria Tzika, Loukas Astrakas, Justin Pearlman, Yuhang Wang, Tilmann Steinberg, Li Shen, Kristen Chambers, and James Ford. Fusing information for tracking tumors. In *The 8th World Congress on Advances in Oncology and 6th International Symposium on Molecular Medicine*, Hersonissos, Crete, Greece, October 16–18, 2003.

[MWS+03] Fillia Makedon, Yuhang Wang, Tilmann Steinberg, Heather Wishart, Andrew Saykin, James Ford, Song Ye, and Li Shen. A system framework for the integration and analysis of multi-modal spatiotemporal data streams: A case study in MS lesion analysis. In *Proceedings of First International IEEE EMBS Conference on Neural Engineering*, pages 495–498, Capri Island, Italy, 2003.

[SCT+01] Li Shen, Ling Cheng, Faye Teng, Fillia Makedon, James Ford, Tilmann Steinberg, and Andrew J. Saykin. A multimedia system for tracing and studying regions-of-interest in brain images. In *IEEE Multimedia Technology and Application Conference*, pages 238–245, Irvine, CA, USA, November 7–9, 2001.

[SFM+03a] Andrew Saykin, Laura Flashman, T. McHugh, C. Pietras, T. W. McAllister, A. C. Mamourian, R. Vidaver, Li Shen, James Ford, Lin Wang, and Fillia Makedon. Principal components analysis of hippocampal shape in schizophrenia. *Schizophrenia Research*, 60(1):206, March 29 – April 2 2003.

[SFM+03b] Li Shen, James Ford, Fillia Makedon, Laura Flashman, and Andrew Saykin. Surface-based morphometric analysis for hippocampal shape in schizophrenia. *NeuroImage*, 19(2):S1004, June 2003.

[SFM+03c] Li Shen, James Ford, Fillia Makedon, Yuhang Wang, Tilmann Steinberg, Song Ye, and A. Saykin. Morphometric analysis of brain structures for improved discrimination. In *MICCAI: Medical Image Computing and Computer Assisted Intervention*, LNCS 2879, pages 513–520, Montreal, Quebec, Canada, November 15–18, 2003.

[SFMS03a] Li Shen, James Ford, Fillia Makedon, and Andrew Saykin. Effective classification of 3D closed surfaces: Application to modeling neuroanatomical structures. In *International Conference on Computer Vision, Pattern Recognition and Image Processing (CVPRIP) in conjunction with Seventh Joint Conference On Information Sciences (JCIS)*, pages 708–711, Cary, North Carolina, USA, September 26–30, 2003. Association for Intelligent Machinery.

[SFMS03b] Li Shen, James Ford, Fillia Makedon, and Andrew Saykin. Hippocampal shape analysis: Surface-based representation and classification. In M. Sonka and J. M. Fitzpatrick, editors, *Medical Imaging 2003: Image Processing*, SPIE Proceedings 5032, pages 253–264, San Diego, CA, USA, February 15–20, 2003.

[SFMS03c] Li Shen, James Ford, Fillia Makedon, and Andrew Saykin. A surface-based approach for classification of 3D neuroanatomic structures. Technical Report TR2003-464, Dartmouth College, Computer Science, Hanover, NH, June 2003.

[SFMS05] Li Shen, James Ford, Fillia Makedon, and Andrew Saykin. A surface-based approach for classification of 3D neuroanatomic structures. *Intelligent Data Analysis, An International Journal*, 8(6), January 2005.

[SFS+01] Andrew J. Saykin, Laura A. Flashman, Li Shen, John Ashburner, Molly Sparling, Annette Donnelly, Fillia Makedon, David Isecke, James C. Ford, Vasileios Megalooikonomou, and Thomas W. McAllister. Hippocampal shape in schizophrenia: A deformation-based morphometric analysis. *NeuroImage*, 13(6):S1096, June 2001.

[SFWM03] T. Steinberg, J. Ford, Y. Wang, and F. Makedon. Similarity searches in heterogeneous feature spaces. In *5th WSEAS Int. Conf. on Mathematical Methods and Computational Techniques in Electrical Engineering (MMACTEE 2003)*, Athens, Greece, 2003.

[SGZ+05] Li Shen, Ling Gao, Zhenwu Zhuang, Ebo DeMuinck, Heng Huang, Fillia Makedon, and Justin Pearlman. An interactive 3d visualization and manipulation tool for effective assessment of angiogenesis and arteriogenesis using computed tomographic angiography. In *SPIE Medical Imaging 2005: Visualization, Image-Guided Procedures, and Display*, San Diego, California, February 12–17, 2005.

[She04] Li Shen. *Three Dimensional Shape Analysis with Parametric Surface Modeling*. PhD thesis, Dartmouth College, Hanover, NH, June 2004.

[SM04a] L. Shen and F. Makedon. Spherical parameterization for 3d surface analysis in volumetric images. In *International Conference on Information Technology (ITCC 2004)*, pages 643–649, Las Vegas, NV, 2004. IEEE Computer Society.

[SM04b] Li Shen and Fillia Makedon. Spherical parameterization for 3D surface analysis in volumetric images. In *ITCC 2004: International Conference on Information Technology*, pages 643–649, Las Vegas, NV, USA, April 5–7, 2004. IEEE Computer Society.

[SM05] Li Shen and Fillia Makedon. Spherical mapping for processing of 3-D closed surfaces. *Image and Vision Computing*, in review, 2005.

[SMS04] Li Shen, Fillia Makedon, and Andrew Saykin. Shape-based discriminative analysis of combined bilateral hippocampi using multiple object alignment. In J. M. Fitzpatrick and M. Sonka, editors, *Medical Imaging 2004: Image Processing*, SPIE Proceedings 5370, pages 274–282, San Diego, CA, USA, February 14–19, 2004.

[SWF+02] Andrew Saykin, Heather Wishart, Laura Flashman, T. W. McAllister, T. McHugh, James Ford, Li Shen, Tilmann Steinberg, and Fillia Makedon. Structure/function relationships in brain disorders: Strategies for mining volume, shape, lesion and bold fMRI activation data. In *Society for Biological Psychiatry Meeting*, Philadelphia, PA, May 16–18, 2002.

[SWM+03] Tilmann Steinberg, Yuhang Wang, Fillia Makedon, Li Shen, Andrew Saykin, and Heather Wishart. A spatio-temporal multi-modal data management and analysis environment for tracking MS lesions. In *SSDBM 2003: 15th International Conference on Scientific and Statistical Database Management*, pages 245–246, Cambridge, MA, July 9–11, 2003. IEEE Computer Society.

[WM03] Yuhang Wang and Fillia Makedon. R-histogram: Quantitative representation of spatialrelations for similarity-based image retrieval. In *The 11th Annual ACM International Conference on Multimedia*, pages 323–326, Berkeley, California, USA, 2003.

[WM04] Yuhang Wang and Fillia Makedon. Application of relief-f feature filtering algorithm to selecting informative genes for cancer classification using microarray data (poster paper). In *Proceedings of the 2004 IEEE Computational Systems Bioinformatics Conference*, Stanford, California, 2004.

[WMC04] Yuhang Wang, Fillia Makedon, and Amit Chakrabarti. R*-histograms: Efficient representation of spatial relations between objects of arbitrary topology. In *Proceedings of the 12th Annual ACM International Conference on Multimedia*, pages 356–359, New York, New York, USA, 2004.

[WMD04] Yuhang Wang, Fillia Makedon, and Robert L. (Scot) Drysdale. Fast algorithms to compute the force histogram. *In submission*, 2004.

[WMF04a] Yuhang Wang, Fillia Makedon, and James Ford. A bipartite graph matching framework for finding correspondences between structural elements in two proteins. In *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2004.

[WMF04b] Yuhang Wang, Fillia Makedon, and James Ford. Mining marker genes for phenotype classification using microarray gene expression data. In *In submission*, 2004.

[WMF+04c] Yuhang Wang, Fillia Makedon, James Ford, Li Shen, and Dina Goldin. Generating fuzzy semantic metadata describing spatial relations from images using the R-histogram. In *Proceedings of the Fourth ACM/IEEE-CS Joint Conference on Digital Libraries*, pages 202–211, Tucson, Arizona, 2004.

[WMFP04] Yuhang Wang, Fillia Makedon, James Ford, and Justin Pearlman. Hykgene: A hybrid approach for selecting marker genes for phenotype classification using microarray gene expression data. *Bioinformatics*, In press, 2004.

[WSM+03] Yuhang Wang, Tilmann Steinberg, Fillia Makedon, Heather Wishart, and Andrew Saykin. Quantifying evolving processes in multimodal 3D medical images. In *The Sixth Annual International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2003)*, pages 101–108, Montréal, Québec, Canada, 2003.

20.3 Contributions

This work has bridged computer science principles with biomedical problems and has contributed to the training of at least 25 graduate students since 1998. It has provided hands-on experience to many undergraduates, who developed interactive tools for analysis and visualization of complex biomedical processes (e.g., the TROI system was developed in an M.S. thesis and later extended twice by undergraduates).

These projects served as a means of transferring solutions to real-world problems into the hands of the clinical and healthcare specialists who introduced their problems. They led to funding for an NSF grant, IDM 0083423 (5/1/2001–4/30/2004) with Makedon as the PI, and have helped attract additional multi-year funding. Several proposals are also pending at the NSF and the NIH.

21 Secure Distributed Resource Sharing (Fillia Makedon)

Research on secure distributed resource sharing in the Dartmouth Experimental Visualization Laboratory (DEVLAB), Fillia Makedon's group, has focused on several areas: negotiation systems, peer-to-peer and sensor network systems, and cyber-security tool development.

21.1 Activities and Findings

1. **Negotiation-based sharing of resources**
   - Developed the Secure Content Exchange Negotiation System (SCENS), a centralized multi-interface access point where humans and agents can index, identify, and set conditions for exchange for resources
   - Formulated the MetaDL framework that underlies the SCENS system, and which provides for the interconnection of existing data repositories through exchanges of metadata
   - Created complementary approaches based on metadata analysis for indexing and retrieval of images and other multimedia objects
   - Developed an automated trust negotiation (ATN) model based on extending current approaches to include third parties
2. **Routing, analysis, and optimization in peer-to-peer and sensor networks**
   - Developed new routing approaches for peer-to-peer and sensor networks
3. **Cyber-security tool development**
   - Created a social network-based tool for analyzing and visualizing activity in peer-to-peer, negotiation, and dynamic situations
   - Developed new collaborative filtering tools based on formation of recommendations from the data in distributed clients
4. **Biweekly educational seminars on topics related to resource sharing, networks, and security**

21.2 Products

A partial list of products includes

1. TecFlow: a tool for visualizing and analyzing network flows in social networks
2. The Information Security Support System (ISSS): a consultation system for cyber-security that links between a digital library of security tools and a security incidents database
3. EcomRisk: an online system for collecting incidents of e-commerce risk in order to support classification, grouping, and other analyses[8]
4. A prototype implementation of SCENS[9]
5. The Global Research and Education in e-Commerce (GREeCOM)[10] online digital library of tools and technical reports related to e-commerce security and computer trends

[8] http://ecomrisk.org
[9] http://scens.cs.dartmouth.edu:8080/SCENS/
[10] http://greecom.org

6. The Electronic Journal for E-Commerce Tools and Applications (eJETA)[11], a peer-reviewed interdisciplinary journal on security and other areas related to technology, policy, and business

Publications

[BFLM03] B. Bhargava, C. Farkas, L. Lilien, and F. Makedon. Trust, privacy, and security: Summary of a workshop breakout session. In *National Science Foundation Information and Data Management (NSF IDM) Workshop*, Seattle, Washington, 2003. Version 2.

[FMS+02] James Ford, Fillia Makedon, Li Shen, Tilmann Steinberg, Andrew Saykin, and Heather Wishart. Evaluation metrics for user-centered ranking of content in metadls. In *Fourth DELOS Workshop on Evaluation of Digital Libraries: Testbeds, Measurements, and Metrics*, Budapest, Hungary, May 6–7, 2002.

[LOFM04] Z Le, Y Ouyang, J. Ford, and F. Makedon. A hierarchical key-insulated signature scheme in the ca trust model. In *Seventh International Conference on Information Security (ISC 2004)*, volume 3225 of *Lecture Notes in Computer Science*, pages 280–291, Palo Alto, CA, 2004. Springer.

[MFS+02] Fillia Makedon, James Ford, Li Shen, Tilmann Steinberg, Andrew Saykin, Heather Wishart, and Sarantos Kapidakis. MetaDL: A digital library of metadata for sensitive or complex research data. In *ECDL 2002: European Conference on Digital Libraries*, LNCS 2458, pages 374–389, Rome, Italy, September 16–18, 2002.

[MKS+03] F. Makedon, S. Kapadakis, T. Steinberg, Song Ye, and L. Shen. Data brokers: Building collections through automated negotiation. Technical Report DEVLAB-SCENS-03-02, Dartmouth College Computer Science Department, March 2003.

[MPCKS03] F. Makedon, G. Pantziou, M. Conalis-Kontos, and C. Sudborough. A safe information sharing framework for e-government communication. In *Electronic Democracy: Information Society and Citizens' Rights*, Athens, Greece, 2003.

[MYS+03] F. Makedon, Song Ye, T. Steinberg, Yan Zhao, Z. Xiao, and B. Sudborough. A security incident sharing and classification system for building trust in cross media enterprises. In *International Conference on Cross-Media Service Delivery (CMSD-2003)*, Greece, 2003.

[MYZ03] Fillia Makedon, Song Ye, and Yan Zhao. On the design and implementation of a web-based negotiation system. In *9th Panhellenic Conference in Informatics (PCI2003)*, Thessaloniki, Greece, 2003.

[MYZ+04] Fillia Makedon, Song Ye, Sheng Zhang, James Ford, Li Shen, and Sarantos Kapidakis. Data brokers: Building collections through automated negotiation. In *SETN'04: The 3rd Hellenic Conference on Artificial Intelligence*, volume 3025 of *LNCS*, pages 13–22, Samos, Greece, May 5–8, 2004.

[OFM+00] C. B. Owen, J. C. Ford, F. Makedon, T. Steinberg, and C. Metaxaki-Kossionides. Parallel text alignment. *International Journal of Digital Libraries*, 3(1):100–114, 2000.

[OM99] C. B. Owen and F. Makedon. *Computed Synchronization for Multimedia Applications*. Kluwer Academic Publishers, 1999.

[11] http://www.ejeta.org

[SCF+00] Li Shen, Ling Cheng, James Ford, Fillia Makedon, Vasileios Megalooikonomou, and Tilmann Steinberg. Mining the most interesting web access associations. In *WebNet 2000–World Conference on the WWW and Internet*, pages 489–494, San Antonio, Texas, USA, October 31 – November 4, 2000.

[SCS+99] Tilmann Steinberg, Ling Cheng, Li Shen, Fillia Makedon, Charles Owen, and Thomas Ottmann. A model for authoring for retrieval (AFR): Retrieval on parallel data streams of recorded lecture information. In *Multimedia Storage and Archiving Systems IV, SPIE Photonics EAST*, pages 423–430, Boston, MA, USA, September 19 – 22, 1999.

[SFO+04] Tilmann Steinberg, James Ford, Yi Ouyang, Li Shen, Yuhang Wang, Wei Zheng, and Fillia Makedon. Tracking resource usage using heterogeneous feature spaces with local exceptions. In *EDBT'04 Workshop: Clustering Information over the Web*, Heraklion–Crete, Greece, March 14, 2004.

[SSCM00] Tilmann Steinberg, Li Shen, Ling Cheng, and Fillia Makedon. Tracking human expression actions in lectures. In *ED-MEDIA 2000: World Conference on Educational Multimedia, Hypermedia & Telecommunication*, Montreal, Quebec, Canada, June 26 – July 1, 2000.

[YBM+02] Song Ye, Matt Bishop, F. Makedon, T. Steinberg, J. C. Ford, L. Shen, and Yuhang Wang. Security concerns in negotiation systems. Technical Report DEVLAB-SCENS-02-02, Dartmouth College Computer Science Department, November 2002.

[YFZM04] S. Ye, J. Ford, S. Zhang, and F. Makedon. Noodle: an ontology description language for automated negotiation. Technical Report DEVLAB-SCENS-04-01, Dartmouth College, 2004.

[YM04] Song Ye and Fillia Makedon. Collaboration-aware peer-to-peer media streaming. In *ACM Multimedia Conference 2004, pp 412-415*, New York City, NY, USA, 2004. ACM Press.

[YMF+02a] Song Ye, F. Makedon, J. C. Ford, L. Shen, T. Steinberg, and Yuhang Wang. An open negotiation system with a web service based implementation. Technical Report DEVLAB-SCENS-02-03, Dartmouth College Computer Science Department, December 2002.

[YMF+02b] Song Ye, F. Makedon, J. C. Ford, L. Shen, T. Steinberg, Yuhang Wang, and S. Kapadakis. A negotiation framework for secure data sharing. Technical Report DEVLAB-SCENS-02-01, Dartmouth College Computer Science Department, October 2002.

[YMF04] S. Ye, F. Makedon, and J. Ford. Collaborative automated trust negotiation in peer-to-peer systems. In *The Fourth IEEE International Conference on Peer-to-Peer Computing (IEEE P2P 2004)*, Zurich, Switzerland, 2004.

[YMS+03] Song Ye, Fillia Makedon, Tilmann Steinberg, Li Shen, Yuhang Wang, Yan Zhao, and James Ford. SCENS: a system for the mediated sharing of sensitive data. In *Proceedings of the Third ACM/IEEE-CS Joint Conference on Digital Libraries*, pages 263–265, Houston, TX, May 27–31, 2003.

[ZOFM03] Wei Zheng, Yi Ouyang, James Ford, and Fillia Makedon. Ontology-based image retrieval. In *6th WSEAS Int. Conf. on Mathematical Methods and Computational Techniques in Electrical Engineering (MMACTEE 2004)*, Athens, Greece, 2003.

[ZYMF04] Sheng Zhang, Song Ye, Fillia Makedon, and James Ford. A hybrid negotiation strategy mechanism in an automated negotiation system. In *ACM Conference on Electronic Commerce (EC 2004)*, New York, NY, 2004.

21.3 Contributions

This work has produced focused, innovative tools for several state of the art problems involving emerging challenges in the global communications environment. It has involved at least 12 graduate students and at least 30 undergraduates in various classes who used and provided feedback on the tools. It has led to support from industry partners and the Institute for Security and Technology Studies (ISTS), as well as funding for two NSF grants with Makedon as the PI: NSF IDM 0308229 (9/1/03–8/31/06) and NSF ITR 0312629 (7/1/03–6/30/06). Several proposals are also pending at the NSF.

22 Simulation and Modeling (David Nicol)

This project looked at problems associated with building software tools, and models, to help quickly analyze communication networks.

With the increasing dependence of national commerce and communications on computer networks, it is very important to study new protocols and network designs in a simulated context, before deploying in an actual setting. Modern communication systems are huge, which means that computer based models of these systems ought also to reflect their size. The most straightforward way of emulating a large network is to build in a model somewhat simplified representations of its components—computer hosts, routing devices, and individually transmitted packets. The very large size of resulting network models creates stresses in (i) holding the simulation model in memory, (ii) running the simulation fast enough to get useful answers in a timely fashion, and (iii) analyzing the results of the simulation.

Our project looks at modeling as one way of dealing with the first two issues. It is part of the larger and very successful SSF project (which is also a success story with its roots in the NSF Networking Special Projects program). SSF built a high-performance parallelized simulation tool-set for simulating very large networks. SSF models typically simulate the transmission of individual packets of information through the network.

Our project build the capability to examine problems in critical infrastructure protection. Early years of the grant built the tools, later years applied them. Specifically we have looked at enhancing the framework to study the effect of the Code Red II and nimda worm traffic on routing infrastructure. This effort seeks to explain why massive routing instability was observed temporally correlated with the worm attacks. The approach has been to model and validate the worm propagation behavior, model the effect worm traffic has on router peformance, and model the effect router performance has on routing protocol (BGP) behavior. Our existing body of BGP modeling code was extended to model dynamic situations where the routers fail, and reboot.

We were able to investigate the hypothesis that worm traffic causes routers to fail, essentially by increasing CPU utilization into failure zones. The randomized probing pattern of worm traffic can do this, because of caching. A packet that carries a new destination, not recently seen by a router, is routed much more slowly (and uses more CPU resources) than subsequent packets. Our modeling experiments have shown that the observed behavior of waves of route withdrawals (in real data) can be explained by the increasing pressure of probing traffic to cause router failure.

In addition to the work on routing, we studied issues in wireless simulation. One example is that we worked with BBN to extend our work on an ad-hoc wireless network simulation testbed, and are using it to study the vulnerability of various routing protocols under consideration.

Our results have very direct application to societal service, both in policy development and the private sector. The Internet is sometimes characterized as the ”Wild Wild West”, in terms of management. It is inevitable that regulatory decisions of some type will led to more organization in the future, and these decision must be well-informed. Simulation modeling gives the capability for making well informed decisions. In the private sector we worked closely with ATT and Motorola, giving them enhanced capability to analyze and manage their operations. We supported large-scale modeling work being done using our tools at Oak Ridge National Laboratory, and Los Alamos National Laboratory.

This project has made extensive use of the SGI Origin 2000 funded by the NSF Infrastructure grant. The large memory, and parallel processing capability are exploited significantly when running large-scale SSF models.

22.1 Activities and Findings

Our goals and objectives for this project were to explore models and software for support of simulation of large complex networks. We focused on routing instabilities caused by adverse traffic, and on developing the simulation infrastructure needed to support that. In the out-of-core context we studied the relationship between pre-fetching and scheduling, extending previous work in the area developed in the context of caching. In the

course of this we've done a good deal of code development, and with experimentation with models to study behavior.

On the educational end, we trained Anna Poplawski, Jason Liu, Srdjan Petrovic, Yougu Yuan, and Guanhua Yan to develop problem identification skills, as well as skills in analysis of literature and presentation of research materials. Liu obtained a Ph.D.

Our essential findings are that caching effects in routers, caused by worm traffic, can explain the routing instability observed in the real internet in the wake of the Code Red II and nimda worm attacks. Future work is focused on studying defensive measures that might be employed by routers, and modifications that might be made to BGP.

22.2 Contributions

Our project contributes to computer science in the system areas of scheduling and memory management, and in the areas of performance and behavior modeling. By its nature and design it contributes to telecommunication engineering; it is an example of how problems in one discipline can motivate research in another. Standing back and looking at what we are trying to accomplish, success will provide capabilities for planning and analysis of communication infrastructure, capabilities that contribute to economic development in the communication sector, as well as policy making in the government sector. It is already beginning to be used in these ways.

22.3 Products

Software. NSF support enabled us to develop and release into the public domain two simulators, DaSSF and SWAN. DaSSF[12] provides a high-performance simulation engine for simulations in areas such as wireline networking, communication system design, and computer architecture. SWAN[13] focuses on ad-hoc wireless network simulation. Both products are used significantly outside of the our immediate research group.

Papers. The following papers were produced by this project during the period of this grant, although some of them have appeared after the end of the grant period. These papers are available at our web site.[14]

Publications

[ANGJ00] Heidi Ammerlahn, David Nicol, Michael Goldsby, and Michael Johnson. A geographically distributed enterprise system. *Future Generation Computer Systems*, 17(2):135–146, October 2000.

[BCM+03] Bill Brown, Andrew Cutts, Dennis McGrath, David M. Nicol, Timothy P. Smith, and Brett Toefel. Simulation of cyberattacks with applications in homeland defense training. In *Proceedings of the AeroSense 2003 Conference*, Orlando, FL, March 2003.

[CN98] Tom Cormen and David Nicol. Performing out-of-core ffts on parallel disk systems. *Parallel Computing*, 24(1):5–20, January 1998.

[CNL+99] James Cowie, David Nicol, Hongbo Liu, Jason Liu, and Andy Ogielski. Towards realistic million-node internet simulations. In *1999 Int'l Conference on Parallel and Distributed Processing Techniques and Applications*, Las Vegas, June 1999.

[CNO99] James Cowie, David Nicol, and Andy Ogielski. Modeling the global internet. *IEEE Computing in Science and Engineering*, 1(1):42–50, Jan.-Feb. 1999.

[12] http://www.crhc.uiuc.edu/~jasonliu/projects/ssf/intro.html
[13] http://www.cs.dartmouth.edu/research/SWAN/
[14] http://www.crhc.uiuc.edu/~nicol//papers-cv/.

[CNT99] Gianfranco Ciardo, David Nicol, and Kishor Trivedi. Simulation of fluid stochastic petri nets. *IEEE Transactions on Software Engineering*, 25(2):207–217, March/April 1999.

[HKLN03] R. Henry, S. Kahan, J. Liu, and David Nicol. An implementation of the ssf scalable simulation framework on the cray mta. In *2003 Conference on Parallel and Distributed Simulation*, San Diego, CA, June 2003.

[LLN03] Micheal Liljenstam, Jason Liu, and David M. Nicol. Development of an internet backbone topology for large-scale network simulations. In *Proceedings of the 2003 Winter Simulation Conference*, pages 694–704, New Orleans, LA, December 2003.

[LN00] Jason Liu and David Nicol. Lock-free scheduling of logical processes in parallel simulation. In *2001 Conference on Parallel and Distributed Simulation*, pages 22–24, Lake Arrowhead, CA, May 2000.

[LN01] Malcolm Low and David Nicol. Consistent modeling of distributed mutual exclusion protocol using optimistic synchronization. In *2001 Conference on Parallel and Distributed Simulation*, pages 137–144, Lake Arrowhead,CA, May 2001.

[LN02] Jason Liu and David Nicol. Lookahead revisited in parallel wireless simulations. In *2002 Conference on Parallel and Distributed Simulation*, pages 79–88, Washington, D.C., May 2002.

[LNBG03] M. Liljenstam, D. M. Nicol, V. Berk, and R. Gray. Simulating realistic network worm traffic for worm warning system design and testing. In *Proceedings of the 2003 Workshop on Rapid Malcode (WORM)*, pages 24–33, Washington, DC, October 2003.

[LNP+01] Jason Liu, David Nicol, Felipe Perrone, Michael Liljenstam, Chip Elliot, and Dave Pearson. Simulation modeling of large-scale ad-hoc sensor networks. In *European Interoperability Workshop 2001*, London England, June 2001.

[LNPL01] Jason Liu, David Nicol, Felipe Perrone, and Michael Liljenstam. Towards high performance modeling of the 802.11 wireless protocol. In *2001 Winter Simulation Conference*, Arlington, VA, December 2001.

[LNPP99] Jason Liu, David Nicol, Brian Premore, and Anna Poplawski. Performance prediction of a parallel simulator. In *1999 Workshop on Parallel and Distributed Simulation (PADS)*, pages 156–164, Atlanta, GA., May 1999.

[LYPN02] M. Liljenstam, Y. Yuan, B. Premore, and D. Nicol. A mixed abstraction level simulation model of large-scale internet worm infestations. In *Proceedings of the Tenth IEEE/ACM Symposium on Modeling, Analysis, and Simulation of Computer Telecommunication Systems*, Fort Worth, TX, October 2002.

[MN03] Weizhen Mao and David M. Nicol. On k-ary n-cubes : Theory and applications. *Discrete Applied Mathematics*, 129(1):171–193, 2003.

[NBF+99] David Nicol, Osman Balci, Richard Fujimoto, Paul Fishwick, Pierre L'Ecuyer, and Roger Smith. Strategic directions in simulation research. In *1999 Winter Simulation Conference*, pages 1509–1520, Phoenix, AZ, December 1999.

[NCL00] David Nicol, James Cowie, and Jason Liu. Safe time-stamps and large scale modeling. In *2000 Workshop on Parallel and Distributed Simulation*, pages 71–78, Bologna, Italy, May 2000.

[NGJ99] David Nicol, Michael Goldsby, and Michael Johnson. Fluid-based simulation of communication networks using ssf. In *1999 SCS European Simulation Conference*, Erlangen, Germany, October 1999.

[Nic98] David Nicol. Scalability, locality, partitioning, and synchronization. In *1998 Workshop on Parallel and Distributed Simulation*, pages 4–11, Banff, Canada, June 1998.

[Nic01] David Nicol. Discrete-event fluid modeling of tcp. In *2001 Winter Simulation Conference*, Arlington, VA, December 2001.

[Nic02a] David Nicol. Analysis of composite synchronization. In *2002 Conference on Parallel and Distributed Simulation*, pages 115–124, Washington, D.C., May 2002.

[Nic02b] David Nicol. Challenges in using simulation to explain global routing instabilities. In *2002 Conference on Grand Challenges in Simulation*, San Antonio, TX, January 2002.

[Nic03a] David M. Nicol. Scalability of garbage collection in java-based discrete-event simulators. In *Proceedings of UKSim 2003*, Cambridge University, England, April 2003.

[Nic03b] David M. Nicol. Scalability of network simulators revisited. In *Proceedings of the Communication Networks and Distributed Systems Modeling and Simulation Conference*, Orlando, FL, February 2003.

[Nic03c] David M. Nicol. Utility analysis of network simulators. *International Journal of Simulation : Systems, Science, and Technology*, 2003.

[Nic03d] David M. Nicol. Utility analysis of parallel simulation. In *2003 Conference on Parallel and Distributed Simulation*, San Diego, CA, June 2003.

[NJY98] David Nicol, Michael Johnson, and Ann Yoshimura. Ides: A java-based distributed simulation engine. In *1998 International Workshop on Modeling Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS)*, pages 233–240, Montreal, Canada, 1998.

[NL01] David Nicol and Jason Liu. Learning not to share. In *2001 Conference on Parallel and Distributed Simulation*, pages 26–55, Lake Arrowhead,CA, May 2001.

[NL02] David Nicol and Jason Liu. Composite synchronization for parallel discrete event simulation. *IEEE Transactions on Parallel and Distributed Systems*, 13(5):433–446, May 2002. To appear.

[NLL03a] D. M. Nicol, M. Liljenstam, and J. Liu. Multiscale modeling and simulation of worm effects on the internet routing infrastructure. In *Proceedings of the Performance Tools 2003 Conference*, Urbana, IL, September 2003.

[NLL03b] David M. Nicol, Jason Liu, and Micheal Liljenstam. Simulation of large-scale networks using ssf. In *Proceedings of the 2003 Winter Simulation Conference*, pages 650–657, New Orleans, LA, December 2003.

[NP00] David Nicol and Felipe Perrone. Cost/benefit analysis of interval jumping in power-control simulation. In *2000 Winter Simulation Conference*, pages 425–431, Orlando, FL, December 2000.

[NPO03] David M. Nicol, Brian Premore, and Andy Ogielski. Using simulation to understand dynamic connectivity at the core of the internet. In *Proceedings of UKSim 2003*, Cambridge University, England, April 2003.

[NY03] David M. Nicol and Guanhua Yan. Discrete-event fluid modeling of tcp background traffic. *ACM TOMACS*, 2003.

[OCN99] Andrew Ogielski, James Cowie, and David Nicol. Modeling 100,000 nodes and beyond: Self-validating and design. In *DARPA/NIST and Workshop on and Validiation of Large-Scale Network and Simulation Models*, Reston, VA, May 1999.

[PN98a] Felipe Perrone and David Nicol. Rapid simulation of wireless systems. In *1998 Workshop on Parallel and Distributed Simulation*, pages 170–177, Banff, Canada, June 1998.

[PN98b] Anna Poplawski and David Nicol. **Nops**: A conservative simulation engine for ted. In *1998 Workshop on Parallel and Distributed Simulation*, pages 180–187, Banff, CA, June 1998.

[PN99] Anna Poplawski and David Nicol. An investigation of out-of-core parallel discrete-event simulation. In *1999 Winter Simulation Conference*, pages 524–530, Phoenix, AZ, December 1999.

[PN00] Felipe Perrone and David Nicol. Using n-body algorithms for interference computation in wireless cellular simulations. In *2000 MASCOTS Conference*, pages 49–56, San Fransisco, CA, August 2000.

[PYN03] Luis Felipe Perrone, Yougu Yuan, and David M. Nicol. Modeling and simulation best practices for wireless ad-hoc networks. In *Proceedings of the 2003 Winter Simulation Conference*, pages 685–693, New Orleans, LA, December 2003.

23 Signal processing (Dan Rockmore)

Professor Rockmore works on Representation Theory, Fast Transforms, Group Theoretic Transforms,Dynamical Systems, Signal Processing and Data Analysis.

23.1 Activities and Findings

Research - the theoretical development and implementation of algorithms for computing spherical harmonic expansions and Fourier expansions on the rotation group.

Major findings: Efficient algorithms for for computing spherical harmonic expansions and Fourier expansions on the rotation group.

Workstations purchased with the award provided a computational environment for algorithm development and implementation.

23.2 Products

Two web sites provide access to our software: *The SOFT Package*,[15] which is software for the computation of FFTs on the rotation group, and *SpharmonicKit*,[16] which is software for the computation of spherical harmonic expansions for functions on the 2-sphere.

Selected papers include:

1. Towards Safe and Effective High-Order Legendre Transforms with Applications to FFTs for the 2-sphere, (w/D. M. Healy Jr. and P. Kostelec), *Advances in Computational Mathematics* **21** (1-2): 59-105, July 2004

2. FFTs for the 2-Sphere – Improvements and Variations (w/D. Healy, P. Kostelec and S. Moore), *J. Fourier Analysis and Appl.* **9** 4: 341–385, July 2003

3. "FFTs for tensor and vector harmonics on the 2-sphere", (with D. Healy, D. Maslen and P. Kostelec), *J. Computational Physics*, **162**, 2000,pp. 514–535.

4. "FFTs on the rotation group", (w/P, Kostelec), Santa Fe Institute Working Paper, 03-11-060, submitted for publication.

5. D. Maslen, M. Orrison and D. Rockmore, "Computing isotypic projections with the Lanczos iteration", *SIAM J. Matrix Analysis and Appplications* 25 (3), 784–803, (2004).

6. R. Foote, G. Mirchandani, and D. Rockmore, "Two-dimensional wreath product transforms", *J. Symbolic Computation* 37 (2) pp. 187–207, (2004).

7. D. Rockmore, "Are you my Mother...Tongue?", *SFI Bulletin*, **19**, no. 1, (2004), pp. 10-15.

8. C. Moore, D. Rockmore and A. Russell "Generic Quantum FFTs", *Proceedings of SODA 2004*, pp. 771-780.

9. C. Moore, D. Rockmore, A. Russell, and L. Schulman "The Hidden Subgroup Problem in Affine Groups: Basis Selection in Fourier Sampling", *Proceedings of SODA 2004*, pp. 1106–1115.

10. D. Rockmore, "Recent Progress and Applications in Group FFTs", *Computational Noncommutative Algebra and Applications*, J. Byrnes, ed., Springer-Verlag (2004), pp. 227–254.

[15] http://www.cs.dartmouth.edu/~geelong/soft
[16] http://www.cs.dartmouth.edu/~geelong/sphere

11. D. M. Healy, Jr., P. Kostelec, and D. Rockmore "Towards Safe and Effective High-Order Legendre Transforms with Applications to FFTs for the 2-sphere", *Advances in Computational Mathematics* 21 (1-2): 59-105, July 2004

12. D. M. Healy, Jr., P. Kostelec, D. Rockmore, and S. Moore "FFTs for the 2-Sphere-Improvements and Variations", *Journal of Fourier Analysis and Applications* 9 4: 341–385, July 2003

13. D. Rockmore, "Network Research: Thinking about the web requires a web of thinking", For SFI End of year report, Spring 2003

14. A. Gamburd J. Lafferty, and D. Rockmore, Eigenvalues spacings for quantized cat maps" *J. Phys A*, **36**, no. 12, (2003) pp. 3487-3499.

15. O. Bastert, D. Rockmore, P. Stadler, and G. Tinhofer, "Landscapes on spaces of trees", *Appl. Math. Comput.*, **131** (2002), no. 2-3, 439–459.

16. W. Hordijk, P. Kostelec, D. Rockmore, and P. Stadler "Fast Fourier transforms for fitness landscapes", to appear in *Applied and Computational Harmonic Analysis*, **12** No. 1, Jan 2002, pp. 57-76.

17. D. Rockmore, "The FFT - An algorithm the whole family can use" *Computing in Science and Engineering*, January/February 2000, Volume 2, Number 1, pp. 60–64.

18. K.-L. Kueh, T. Olson, D. Rockmore and K.-S. Tan, "Nonlinear approximation theory on finite groups", Technical Report PMA-TR99-191, Department of Mathematics, Dartmouth College, November, 1999, *J. Fourier Analysis and Applications*, **7**, no. 3, (2001) pp. 257-281.

19. R. Foote, G. Mirchandani, D. Rockmore, D. Healy and T. Olson "A wreath product group approach to signal and image processing: Part I – multiresolution analysis", *IEEE Trans. in Signal Processing*, vol. 48(1), 2000, pp. 102–132

20. R. Foote, G. Mirchandani, D. Rockmore, D. Healy and T. Olson "A wreath product group approach to signal and image processing: Part II – convolution, correlations and applications", *IEEE Trans. in Signal Processing*, vol. 48(3), 2000, pp. 749–767.

21. D. Maslen and D. Rockmore, "Double coset decompositions and computational harmonic analysis on groups," *J. Fourier Analysis and Applications*, Vol. 6(4), 2000. pp. 349-388.

23.3 Contributions

New algorithms and software for computing Fourier transforms in new geometries (i.e., spaces other than R^n) were created and implemented.

The software is being used by researchers in 3D database search, computer vision, climate modeling, and atmospheric sciences.

The software is freely available so can be used for all forms of research and education as personnel see fit.

Human resources. Several students were trained as part of this research.

- Senthil Periaswamy, Ph.D., CS
- Karolyn Abrams, Honors Thesis
- Douglas Warner, Ph.D., Math
- Michael Orrison, Ph.D., Math
- Rob Savell, Ph.D. (expected), CS

24 Robotics (Daniela Rus)

24.1 Activities

In the course of this project, we have designed, built, and used macro and MEMS scale robots to study distributed robotics, with the goal of creating more versatile robots by developing reconfigurable massively-parallel robot systems. Reconfiguration can be thought of broadly as the property of an uncoupled distributed system of robots that can redefine their cooperative roles for solving a task, for example distributed manipulation. A different view of reconfiguration considers connected distributed systems of robots that can physically change shape without human intervention. We have examined aspects of both domains by working on distributed manipulation, MEMS scale locomotion, and self-reconfiguring robots. We have also examined applications to graphics and computational biology.

24.1.1 Self-reconfiguring Robots

A robot designed for a single purpose can perform some specific task very well, but it will perform poorly on a different task, in a different environment. This is acceptable if the environment is structured; however if the task is in an unknown environment, then a robot with the ability to change its shape to suit the environment and the required functionality will be more likely to succeed than a fixed-architecture robot. The goal of this research direction is to develop high degree of freedom self-reconfiguring robots. Self-reconfiguring robots have the ability to adapt to the operating environment and the required functionality by changing shape. They consist of a set of identical robotic modules that can autonomously and dynamically change their internal geometric structure for locomotion, manipulation, or sensing purposes, in order to optimally carry out a variety of tasks. For example, a self-reconfiguring robot system could self-organize as a snake shape to traverse a narrow tunnel and reorganize as a multi-legged walker upon exit to traverse rough terrain.

We focused on unit modular robots, in which all modules are identical. The geometric, mechanic, and kinematic structure of a unit-modular system define its specific approach to self-reconfiguration. In our IROS and Autonomous Robots papers we examined a theoretical basis for self-reconfiguration and proved that a system composed of a specific unit module is self-reconfigurable if it satisfies two properties: (1) groups of unit modules can be assembled into arbitrarily shaped rigid structures and (2) in any structure composed of unit modules, some unit module can be relocated arbitrarily on the structure without human intervention. The first property ensures that any geometric structure can be aggregated from some collection of modules. The second property provides for shape metamorphosis in a general way: given a starting structure S and a goal structure G, it makes it possible to construct G from S incrementally. This theoretical basis for designing self-reconfiguring robots guided us to two new robot designs and self-reconfiguration planning algorithms.

Using this design principle we designed and built two self-reconfiguring robot systems.

The Robotic Molecule. The Robotic Molecule is a four-degree-of-freedom, small scale module that can aggregate with other identical modules to form three-dimensional dynamic structures. The Robotic molecule has been the first module capable of self-reconfiguration in three dimensions. The goal for our design was to create a solid shape that could closely be packed in a three dimensional space and that could move by attaching itself to similar units.

A robotic molecule can connect to other identical modules to create dynamic three-dimensional structures. It is possible to approximate arbitrary three-dimensional objects by packing robotic molecules. More interesting questions concern how to metamorphose a given structure into a desired structure and how to use self-reconfiguration for locomotion and manipulation. These questions can be all formulated as motion planning problems. We addressed this challenge in two parts. First, we developed a set of device-level primitives for controlling the motion of one module relative to a structure of modules. Second, we developed planning algorithms that use the motion primitives to compile automatically locomotion gaits using self-reconfiguration. The key observation for planning is that our self-reconfiguring systems consist of identical modules. Since all

the modules are identical and interchangeable, it is not necessary to compute goal locations for each element. Thus, self-reconfiguration is different than the related warehouse problem (where modules are assigned unique ids and have to be placed at desired locations), which is intractable. Unlike the case of fixed-architecture robots, planning for locomotion also requires dynamic considerations. Self-reconfiguring robots move under gravity and planners have to create stable structures at all times (not just for the initial and final configuration).

An on-line greedy algorithm allows a group of molecules use self-reconfiguration for stable locomotion. The robot moves straight to the goal, by climbing on top of any obstacles along the way (if the number of modules is sufficient to reach the top of the obstacle). Our work in locomotion planning has also resulted in bounds for the minimum number of robotic molecules necessary for translations, rotations, stacking, and stair climbing.

We have built a 4 Molecule Robot system and implemented locomotion algorithms on this system. This work was reported in Robots and Autonomous Systems and in the Communications of the ACM.

The Crystal robot. The Crystalline module has square (cubic in three dimensions) shape with connectors to other modules in the middle of each face. It is activated by three binary actuators that permit the side length of the square to shrink and expand by a factor of two and to make or break connections. This actuation scheme allows an individual module to relocate to arbitrary positions on the surface of a structure in constant time. Previous systems necessitate linear time in the number of modules on the surface. The expansion/compression actuation of the Crystalline Atom allows a module to relocate from point A to point B by traveling through the volume of the structure. We developed and analyzed an algorithm called melt-grow for self-reconfiguration where the initial and the goal structures have the same volume. The melt-grow planner achieves shape morphing by using an intermediate structure, which is a projection of the robot modules into a pool on the ground. This approach leads to reconfiguration algorithms that maintain stability in real-world environments where gravity is present. The planner runs in O(n^2) time, where n is the number of modules in the Crystal. This algorithm shaves an O(n) factor from all previous self-reconfiguration planning algorithms. In our lab, we have built ten Crystalline Atoms and experimented with using self-reconfiguration for locomotion and for shape morphing tasks. This first system and supporting algorithms are centralized.

We have built two generations of Crystal robot systems. The first generation consisted of 10 modules. The second generation Crystal system consists of 16 modules that have improved actuation and sensing, point to point communication, and an additional degree of freedom. It also has a software infrastructure for distributed algorithms. We have implemented several distributed locomotion, splitting, and merging algorithms on this system. This work was reported in IROS 2000 and Mechatronics.

Distributed Algorithms. Early work involved distributed reconfiguration planning for 2D unit-compressible systems, which was later extended to both 3D unit-compressible systems and 3D surface-moving systems. We developed the PacMan algorithm, a technique for distributed actuation and planning. The basic concept behind PacMan is that the modules plan paths for themselves in a distributed fashion. These paths can then be actuated in an asynchronous fashion, so that local motion control is sufficient for reconfiguration. Both planning and actuation are covered under PacMan. We developed two versions of the algorithm along with correctness analysis and show the parallel actuation capability of the algorithm. Namely, for the reachable class of shapes, any reconfiguration can be planned, and any set of paths that does not contain a cycle can be actuated in parallel without synchronization. Finally, we have also extended PacMan to an abstract model of a self-reconfiguring system that does not use compressible modules, to illustrate how this type of algorithm may be used in a more general context. This work was reported in IROS 2000, WAFR 2002, and has been submitted to the International Journal on Robotics Research.

Generic Algorithms. Significant work was also performed in the area of distributed control for a generic abstract model of self-reconfiguring systems. In this work, we attempted to develop algorithms for an abstract model of a self-reconfiguring module. In this way, the algorithms can be fairly simple and provable as well as

extensible to a variety of modular robot hardware. The algorithms developed are based on cellular automata, using geometric rules to control module actions. The actuation model used is a general one, presuming that modules can generally move over the surface of a group of modules. These algorithms can then be instantiated on to a variety of particular systems. Correctness proofs of the rule sets are also given for the generic geometry, with the intent that this analysis can carry over to the instantiated algorithms to provide different systems with correct locomotion algorithms.

The first algorithms developed in this framework were for locomotion. The locomotion rule sets allow a group of modules to move forward without specifying any intermediate shapes or using any global information. These algorithms work both for planar and 3D modules, and can be shown to always generate forward motion without deadlocks. An extension to this work that enabled division and recombination of groups in addition to simple locomotion was also developed. Since a group of self-reconfiguring modules can physically disconnect into several smaller groups to perform operations (such as exploration) in parallel, algorithmic support for this operation is desirable. Similar rule-based algorithms allow division and locomotion in two and three dimensional systems as well as recombination in two dimensions. Correctness analyses for these extensions have also been presented.

In a similar vein, we developed distributed algorithms with which a self-reconfiguring robot can determine its shape. In particular, we are interested in the goal recognition problem — that is, given a particular goal shape, can the modules determine whether they have achieved the shape. The algorithms we developed work for both 2D and 3D lattice-based systems, and use only local messaging to make the decisions. Messages are passed around the perimeter of the group (or through the body in 3D) and the current shape is locally compared to the goal shape in each position in such a way that a global decision is correctly made. In addition, the 2D algorithm requires no memory and only O(n) time for n modules, whereas in 3D O(n) memory is required, but still only O(n) time, since all modules compute in parallel.

Finally, we have also instantiated several algorithms onto the Crystalline Atomic robot, a self-reconfiguring robot under development in our laboratory. Specifically, the generic division algorithm mentioned above was implemented directly, as was our goal recognition algorithm. In addition, distributed locomotion algorithms designed specifically for unit-compressible actuation were written and instantiated. The Crystal robot is unique among self-reconfigurable systems in that the modules are completely untethered during operation, so implementation necessitates complete distribution of the algorithms as well as the ability for modules to communicate autonomously among themselves. To support this work, we developed a software and communication infrastructure for the modules. We were able to perform experiments that empirically verified that the theoretical algorithms can be made to work on a physical self-reconfiguring system in a straightforward fashion.

This work was reported in ICRA 2002, DARS 2002, Mechatronics, and the International Journal of Robotics Research.

Heterogeneous Self-reconfiguration Planning. The majority of work on planning for self-reconfiguring robots concerns homogeneous systems, where all modules are identical and interchangeable. We considered the planning problem for heterogeneous systems that include specialized modules—for example camera modules that need to stay at the top of the robot, special wheel modules, battery modules, or any other specialization that will map into position constraints on the location of the special modules. We considered the relationship between this problem and the warehouse problem which is intractable. We found that the amount of free space around the robot is the key resource that allows polynomial time solutions to the problem. We developed several centralized and distributed algorithms for morphing heterogeneous self-reconfiguring robots and showed that they have the same asymptotic complexity and the homogeneous case. More specifically, the first algorithm examined self-reconfiguration when the robot has as much free space to extend as needed (what is needed is upper bounded by the number of modules in the robot). The second algorithm examined self-reconfiguration when only a small crust of free space is available around the robot. The third algorithm considers position constrains within this crust and can be translated into an algorithm for locomotion by self-reconfiguration within obstacles. All these algorithms have been developed as centralized and decentralized algorithms. They have been proved

correct and analyzed for the computation complexity. They have been implemented in simulation and restricted versions have been implemented on the Crystal robot. This work has been reported in IROS 2003, ICRA 2004, and DARS 2004.

24.1.2 MEMS Robots

Untethered Scratch Drive Actuators The goal of this research is to enable autonomous locomotion at the micro-scale. To achieve this with existing microfabrication technologies, we are investigating the development of scratch drive actuators that can operate in an untethered fashion. We are fabricating these devices using the MUMPS (Multi-User MEMS Process) fabrication process from Cronos Integrated Microsystems (formerly MCNC).

Our untethered scratch drives comprise two novel systems that are currently under development. The first is a capacitive power couple for delivering power to devices that are not physically wired to the substrate. This will allow the devices to locomote in an untethered fashion. The second novel system in our untethered scratch drives is the self-release mechanism by which our devices cut their ties with the fabrication substrate.

In order to fabricate these devices, we need a layer of insulation between two layers of conductive silicon. The MUMPS process does not provide an insulating layer that will work in this regard. So, we are developing ways of post-processing MUMPS devices so as to introduce an intermediate layer of insulation after the fact.

A scratch drive is a direct-drive actuator that operates through electrostatic attraction. It is composed of a thin silicon plate with a bushing at the front end. The plate is typically in the range of 80 microns long and wide, and 1.4 microns thick. The bushing height is typically in the 1-2 micron range. When a voltage is applied between the silicon plate and the substrate beneath it, the plate is drawn down into contact with the substrate. Since the front of the plate is supported by the bushing, strain energy is stored in the plate, and the edge of the bushing is pushed forwards. When the voltage is removed, the strain is released and the scratch drive plate moves forwards.

When an AC signal is applied, the above cycle is continuously repeated, and the scratch drive moves forward in a step-wise manner. Scratch drives have been operated at frequencies as low as 10 Hz, and as high as 2 kHz, with voltages ranging from 30 - 200 V.

The scratch drives that we have built does not require a direct electrical connection (wire) in order to apply the drive voltage. These devices receive power from an underlying grid of electrodes that do not restrict the movement of the drive. This enables us to study the locomotion of the drive when it is not guided by physical tethers or rails.

The electrodes are arranged such that given any two adjacent electrodes, one has positive voltage and the other has negative voltage. Each electrode is smaller than a scratch drive, so that no matter the position or orientation of a given scratch drive, it always lies above some area of positive voltage, and some area of negative voltage.

Because the scratch drive is made of conductive silicon but is coated with an insulating layer, charge flows within the scratch drive in response to the voltage on the underlying electrodes, but it will not flow into the electrodes. So, there is charge build-up on the underside of the scratch drive plate. This charge build-up causes the electrostatic attraction that results in motion of the scratch drive.

The typical drive voltage of a scratch drive actuator is 150 V In order to achieve an equivalent voltage with the capacitive couple, approximately 300 volts must be applied between the high-voltage electrodes and the grounded electrodes.

The scratch drive devices must remain attached to their underlying substrate throughout the fabrication process. However, they must be detached from the substrate prior to operation. So, a release mechanism is required. We would like this release mechanism to be compatible with batch fabrication. i.e. the devices should be able to self-release.

To accomplish this, the scratch drives are originally suspended above the substrate by a long thin beam. The beam is intentionally flawed where it joins the scratch drive plate. When voltage is first applied to the

electrodes beneath the scratch drive, the electrostatic attraction draws the device towards the substrate. The resulting deformation stress in the beam is concentrated at the intentional flaw. The beam then snaps at the flaw and the device is released.

To create the untethered scratch drive actuator we developed a procedure to post-process MUMPS chips. MUMPS is a three-layer polysilicon fabrication process with a top layer of metal. (The metal layer is incompatible with our post-processing steps, and is therefore not used in our devices.) The silicon layers are shaped and separated by intervening sacrificial layers of silicon dioxide. The oxide layers are removed with a hydrofluoric acid wet etch.

In our devices, the sacrificial oxide separates the scratch drives from the underlying electrodes. When the oxide is removed, the drives are suspended about 2 microns above the electrodes. We need to add a layer of insulator that is thick enough to prevent dielectric breakdown at the driving voltage of approximately 200 V, but that is not so thick that it fuses the scratch drives to the electrodes.

To do this, we grow thermal oxide on the silicon by heating in water vapor at 700° C. (The relatively low oxidation temperature prevents thermal stress from deforming the thin scratch drive plates.) Since this process coats all silicon surfaces in insulator, there is then no way to make electrical contact with the substrate. So, the devices are lithographically patterned with a very low-resolution mask, and holes are opened up in the oxide above electrical contact pads.

This work has been reported in IEEE MEMS 2003 and in JMEMS.

24.2 Products

This project has produced many papers, in addition to working prototypes of many of the robots.

Publications

[BBR01] Z. Butler, S. Byrnes, and D. Rus. Distributed motion planning for modular robots with unit-compressible modules. In *Proceedings of the International Conference on Intelligent Robots and Systems (IROS)*, 2001.

[BFR02a] Z. Butler, R. Fitch, and D. Rus. Distributed control for unit-compressible robots: Goal-recogition, locomotion and splitting. *IEEE/ASME Trans. on Mechatronics*, 7(4):418–30, December 2002.

[BFR02b] Z. Butler, R. Fitch, and D. Rus. Distributed goal recognition algorithms for modular robots. In *Proceedings of IEEE ICRA*, 2002.

[BFR02c] Z. Butler, R. Fitch, and D. Rus. Experiments in distributed control of modular robots. In *Experimental Robotics VIII*, pages 307–16, 2002.

[BFR02d] Z. Butler, R. Fitch, and D. Rus. Experiments in distributed locomotion with a unit-compressible robot. In *Proceedings of the International Conference on Intelligent Robots and Systems*, pages 2813–8, 2002.

[BMR02] Z. Butler, S. Murata, and D. Rus. Distributed replication algorithms for self-reconfiguring modular robots. In *Proceedings of Distributed Autonomous Robotic Systems 5*, 2002.

[BR02] Z. Butler and D. Rus. Distributed motion planning for 3-D robots with unit-compressible modules. In *Workshop on the Algorithmic Foundation of Robotics*, 2002.

[DGR99] B. Donald, L. Gariepy, and D. Rus. Experiments in constrained prehensile manipulation: distributed manipulation with ropes. In P. Corke, editor, *Experimental Robotics VI*. Springer Verlag, 1999.

[DGR00] B. Donald, L. Gariepy, and D. Rus. Distributed manipulation of multiple objects with ropes. *Proceedings of the IEEE International Conference on Robotics and Automation*, 2000.

[FBR03] R. Fitch, Z. Butler, and D. Rus. Reconfiguration planning for heterogeneous self-reconfiguring robots. In *Proceedings of the International Conference on Intelligent Robots and Systems*, pages 1813–18, 2003.

[FRV00] R. Fitch, D. Rus, and M. Vona. A basis for self-repair robots using self-reconfiguring crystal modules. In *Proceedings of Intelligent Autonomous Systems 6*, 2000.

[KR98] Keith Kotay and Daniela Rus. Motion synthesis for the self-reconfiguring robotic molecule. In *Proceedings of the International Conference on Intelligent Robots and Systems*, 1998.

[KR99] K. Kotay and D. Rus. Locomotion versatility through self-reconfiguration. *Robotics and Autonomous Systems*, 26:217–232, 1999.

[KR00a] K. Kotay and D. Rus. Algorithms for self-reconfiguring molecule motion planning. In *Proceedings of the International Conference on Intelligent Robots and Systems*, pages 2184–93, 2000.

[KR00b] K. Kotay and D. Rus. Scalable parallel algorithm for configuration planning for self-reconfiguring robots. In *Proceedings of the Society of Photo-Optical Instrumentation Engineers*, Boston, 2000.

[KRVM98] Keith Kotay, Daniela Rus, Marsette Vona, and Craig McGray. The self-reconfiguring robotic molecule: design and control algorithms. In *Proceedings of the Workshop on Algorithmic Foundations of Robotics*, 1998.

[PR02] Ron Peterson and Daniela Rus. Interacting with sensor network. In *Proceedings of the Australian Conference on Robotics and Automation*, Auckland, NZ, November 2002.

[RBKV02] D. Rus, Z. Butler, K. Kotay, and M. Vona. Self-reconfiguring robots. *Communications of the ACM*, 45(3), March 2002.

[RV99] D. Rus and M. Vona. Self-reconfiguration planning with unit compressible modules. In *Proceedings of IEEE ICRA*, pages 2513–20, 1999.

[RV00] D. Rus and M. Vona. A physical implementation of the crystalline robot. In *Proceedings of IEEE ICRA*, 2000.

[RV01] D. Rus and M. Vona. Crystalline robots: Self-reconfiguration with unit-compressible modules. *Autonomous Robots*, 10(1):107–24, 2001.

[SRK03] Sanjiv Singh, Daniela Rus, and Vijay Kumar, editors. *Workshop on Robotics in Emergency Response*, Las Vegas, October 2003. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).

24.3 Contributions

The contributions of this work are many and varied.

- Our work has demonstrated that we can develop scalable distributed controllers for modular self-reconfiguring robots. Our controllers are distributed, based on local information only, and provably correct. Some of our algorithms are architecture-specific. Others are generic in that they can be instantiated to many different types of robot architectures. We have shown distributed algorithms of this flavor to

several tasks: (1) shape morphing; (2) locomotion; (3) division of a large robot into smaller robots; (4) merging of several small robots into a large robot; (5) self-repair and (6) goal recognition.

Our work has also shows that for self-reconfiguration planning homogeneity is not that different than heterogeneity computationally (that is, the asymptotic omputation time for the plan is the same but the constants are different.) The work has also shown a close replationship between heterogeneous planning and sorting and the warehouse problem. These results lead to guiding principles for designing and building the next generation of self-reconfiguring robots.

The implementation of these algorithms on the Crystal robot has shown that it is possible to develop an autonomous robot capable of the above tasks, when the tasks are controlled in a distributed way. Our robots performed continuously for 5 days at SIGGRAPH 02 and for three additional days at AAAI 02. These demonstrations have shown that our hardware is robust and the software is efficient.

Our experiments have also pointed out some weaknesses of the Crystal hardware. The most important improvement should be in the connector design. Further improvements in the communication infrastructure and power supply would also be useful.

Our work has also demonstrated new capabilities for MEMS actuators. We have developed a robust process for post-processing MUMPS and used this process to build the first untethered scratch drive actuators. Our extensive experimentation has demonstrated that the scratch drive actuators are reliable and can move very fast.

- This project led to the creation of the first self-reconfiguring robot with unit-compressible modules. We developed the hardware and the software infrastructure for controlling this robot in a distributed fashion. We developed several distributed algorithms for reconfiguration planning for this robot. We proved the correctness of, and implemented, these algorithms. We demonstrated the resulting systems at SIGGRAPH 2002 and AAAI 2002.

- We started a collaboration with Prof. Duane Compton from the Dartmouth Medical School to apply the rule-based approach to local control for self-reconfiguring robots to spindle formation in cell mitosis. The process of cell mitosis is similar to the processes that lead to self-organization in several ways. In cell mitosis, a collection of rods called microtubules are organized as a spindle by the action of three different types of protein motors. This cell-level self-organization is driven by local interactions and brownian motion. We are collaborating to develop a rule-based model like those developed in our papers (ICRA 2002, DARS 2002, and IROS 2002) to this biological problem.

- One of the main research topics of this project is in support of self-organization and our work reaches beyond the robotics community Self-organization will not only lead to more capable robots, but also to a better understanding of biological systems, which are self-organizing at meny different levels, starting with the cell. Clearly living self-organizing systems are able to perform complex tasks even when built of simple components with simple rules, and we believe these can inform our development of modular robotic systems. Development of self-organizing robots can also help in our understanding of living systems in that they are more easily controlled and altered while exhibiting similar organizational behavior. Thus, our quest to develop self-organizing robots can be viewed as a computational approach to understanding self-organization, where the resources include hardware (sensors and actuators), state (internal and external), communication, and computation.

  Self-organization is a key attribute to understanding biological systems and ultimately designing machines with higher cognition capabilities.

Human resources. Numerous graduate and undergraduate students, postdocs, and others worked in the Dartmouth Robotics Lab throughout the period.

- Undergraduates: Sean Byrnes (now at Cornell) Fred Reiss (now at U.C. Berkeley, NSF fellowship), Marty Vona (now at MIT, NSF fellowship), Christine Alvarado (now at MIT, NSF fellowship), Michael Ross (now at MIT, NSF fellowship), Michael Taylor (now at MIT), Ken Yasuhara (now at U. Washington), David Gondek(now at Brown), Dawn Lawrie (PhD from UMass, assistant professor of CS at Loyola Univ. in Baltimore MD), Michael Shin (now at JHU), Miranda Barrows (now at Umass), Alik Widge (now in MD/PhD program at Pitt and CMU), Marisa Kolodny (architecture at MIT? starting in 2003), Michael Brewer (now at MIT in math), Greg Friedland (now at UCSF in biology) Peter deSantis, Scott Silver, Ahsan Kabir, Joe Edelman, Jason Kochel, Robert Leathern, Mathew Saldo, Morgan Soutter, David Hoffer, Vishesh Khemani, Damon Smith, Cem Paya, Hans Kieserman, Bill Bleier, Tarim Wasim, Michael Carr, Tom Millet, Josh Mills, Jun Shen, David Zipkin (at MIT), Mintcho Petkov, Michael Pryor, Erik White, Jeffrey Zimpleman, Matt Carter, Andrew Ferrone, Jeffrey Steeves, David Black-Schaffer (at Stanford) , Anne Loomis, Sarah Honorowski, David Marmaros, Murphy Stein, Jayson Farrell, Frederick Strathmeyer, Michael deRosa (at CMU), Tom Temple (now at MIT).

- Graduate students: The equipment we developed and purchased as part of this project has also lead to the training of several PhD students: Keith Kotay, Craig McGray, Robert Fitch, Igor Paprotny, Fred Henle, Qun Li, and Jon Howell. Of these students several have received their PhD degrees. Keith Kotay is now a postdoc at MIT. Rob Fitch will be joining NICTA in Sydney this Fall. Qun Li is a tenure track assistant professor at College of William and Mary. Jon Howell is a researcher at Microsoft Research. Craig McGray has yet to graduate but already has an offer for a postdoc position at NIST.

- Postdoc: We have trained one postdoc in research and teaching as part of this project. Zack Butler (now on the faculty at Rochester Institute of Technology) participated in research meetings to define a research agenda, research meetings to solve problems and package them for publication, and research meetings to define experiments. He also participated in several professional meetings and presented papers at IROS00, IROS01, IROS02, ICRA02, DARS02, ISER02, WAFR 2002, IROS03, and ICRA04. He also co-authored 3 journal papers, and was a member on one PhD committee. He was also active in proposal preparation and has taught 2 courses.

- We hosted William (Bill) Church, a local highschool teacher of physics in our lab and worked with him on understanding the physics of Crystal locomotion. Our postdoc Zack Butler helped to train Bill to use our robots. The experience gave Bill Church the experience of formulating and approaching research problems. Bill was inspired to write new proposals for working outside of his classroom.

25 Sensor Networks (Daniela Rus)

Daniela Rus and her group were active in sensor networking, producing numerous algorithmic and experimental results in the field.

25.1 Activities and Findings

Distributed adaptive sensor networks are reactive computing systems, well-suited for tasks in extreme environments, especially when the environmental model and the task specifications are uncertain and the system has to adapt to it. A collection of active sensor networks can follow the movement of a source to be tracked, for example a moving vehicle, it can guide the movement of an object on the ground, (for example a surveillance robot,) or it can focus attention over a specific area, (for example in the case of a fire we would like the sensor network to localize its source and track its spread.) Our goal is to design, build, analyze, and test reactive sensor networks that employ the principles of self-organization and self-assembly to combine structure with function. To this end, we have built a physical sensor network consisting of 50 Mote nodes and algorithmic routing, planning and control support that will allow its units to dynamically and automatically configure themselves into a variety of network configurations in response to environmental changes and in support of different tasks.

Sensors detect information about the area they cover. They can store this information locally or forward it to a base station for further analysis and use. Sensors can also use communication to integrate their sensed values with the rest of the sensor landscape. Users of the network (robots or people) can use this information as they traverse the network. We illustrate this property of a reactive sensor network in the context of a guiding task, where a moving object is guided across the network along a safe path, away from the type of danger that can be detected by the sensors.

The guiding application can be formulated as a robotics motion planning problem in the presence of obstacles. The interesting areas of the sensor network are those where sensors have triggered. They can be represented as obstacles. Such areas may include excessive heat (from volcanoes, fire, etc), people, etc. We assume that each sensor can sense the presence or absence of such an event. An event configuration protocol run across all the nodes of the network creates the event map. We do not envision that the network will create an accurate geometric map, distributed across all the nodes. Instead, we wish for the nodes in the network to provide some information about how far from the event each node is. If the sensors are uniformly distributed, *the smallest number of communication hops to a sensor that triggers "yes" to the event* is a measure of the distance. The goal is to find a path for the moving object that moves toward the events (or avoids them, depending on the application.) The user may ask the network regularly for where to go next. The nodes within broadcasting range from the user supply the next best step.

Inspired by robotics motion planning, we developed several protocols for the distributed guidance problem across sensor networks and reported the details of these algorithms in a paper that has just been accepteed by Mobicom 2003. The map can be constructed incrementally and adaptively as an artificial potential field using hop-by-hop communication. The "obstacles" correspond to events and have repulsing values and the goal has an attracting value. The potential field is computed in the following way. Each node whose sensor triggers "event" diffuses the information about the event to its neighbors in a message that includes its source node id, the potential value, and the number of hops from the source of the message to the current node. This message is used to update the potential value at the current node. The node then broadcasts a message with its new potential value and number of hops to its neighbors.

The potential field information stored at each node can be used to guide an object equipped with a sensor that can talk to the network in an on-line fashion. The safest path to the goal can be identified with a distributed protocol using dynamic programming. In our Mobicom 2003 paper we prove that our algorithm does not get stuck in local minima. A user of the sensor network can get continuous feedback from the network on how to traverse the area. The user asks the network for where to go next. The neighboring nodes reply with their current values. The user sensor chooses the best possibility from the returned values.

The navigation guidance application is an example of how simple nodes distributed over a large geographical area can assist with global tasks. This application relies on the ability of the network user to interact with the network as a whole and with specific nodes in the network. This interaction is directed at retrieving data from the network (such as collecting local information from individual nodes and collecting global maps from the network) and injecting data into the network (such as configuring the network with a new task or reprogramming its nodes).

The ability to re-task and reposition sensors in a network by sending state changes or uploading new code greatly enhances the utility of such a network. It allows different parts of the network to be tailored to specific tasks, capabilities to be added or changed, and information to be stored in the nodes in the network. When robots or people interact with the network, the sensors become an extension of the user capabilities, basically extending their sensory systems and ability to act over a much large range.

We have developed and built a hand-held device that allows a user of the network (a human or a robot) to interact with the network as a whole or to talk to individual nodes in the network. This device is called a *sensory Flashlight* and is based on the optical flashlight metaphor. When pointed in a specific direction, the Flashlight collects information from all the sensors located in that direction and provides its user with sensory feedback. The device can also issue commands to the sensors in that direction.

Applications of the Flashlight device for interacting with the sensor network include: (1) Guiding robots or people along paths that may change over time; (2) Reconfiguring a wireless sensor network in a patterned way; (3) Interacting with a wireless sensor network, both consuming and providing information stored within the network, changing and reacting to its topology, re-tasking the network; (4) Invisible markup of a geographic region with information; (5) Sensor management; and (6) Efficiency improvements in message routing.

We have examined the application of these resulting routing algorithms in the context of a tracking task and presented our results in a paper recently accepted by the 2003 ACM SensSys conference. More specifically, we investigated the computational power of sensor networks in the context of a tracking application by taking a minimalist approach focused on binary sensors. The binary model assumption is that each sensor network node has sensors that can detect one bit of information and broadcast this bit to a base station. We examine the scenario in which the sensor's bit is whether an object is approaching it or moving away from it. We analyzed this minimalist binary sensor network in the context of a tracking application and show that it is possible to derive analytical constraints on the movement of the object and derive a tracking algorithm. We also showed that a binary sensor network in which sensors have only one bit of information (whether the object they sense is approaching or moving away) will give accurate predictions about the direction of motion of the object but do not have enough information content to identify the exact object location. For many applications predicting directional information is enough—for example in tracking a flock of birds, a school of fish, or a vehicle convoy. However, it is possible to pin down the exact location by adding a second binary sensor to each node in the net. If we include a proximity sensor that allows each node to report detecting the object in its immediate neighborhood we can determine the direction and location of the moving target.

This minimalist approach to sensor networks gives us insight into the information content of the tracking application, because it gleans the important resources for solving this task. By studying minimalist sensor networks we learn that the binary sensor network model with one bit gives reliable direction information for tracking, but an additional bit provided by a proximity sensor is necessary to pin down exactly the object location.

Our tracking algorithms have the flavor of particle filtering and make three assumptions. First, the sensors across a region can sense the target approaching or moving away. The range of the sensors defines the size of this region which is where the active computation of the sensor network takes place (although the sensor network may extend over a larger area.) The second assumption is that the bit of information from each sensor is available in a centralized repository for processing. This assumption can be addressed by using a simple broadcast protocol in which the nodes sensing the target send their id and data bit to a base station for processing. Because the data is a single bit (rather than a complex image taken by a camera) sending this information to the base station is feasible. Our proposed approach is most practical for application where the target's velocity is slower than the data flow in the network, so that each bit can actually be used in predictions. However, since the accuracy of our

trajectory computation depends on the number of data points, the predictions are not affected by the velocity of the target relative to the speed of communication. The third assumption is that an additional sensor that supplies proximity information as a single bit is available. Such a sensor may be implemented as an IR sensor with thresholding that depends on the desired proximity range, and can also be derived from the same basic sensing element that provides the original direction bit of information.

The tracking application and many other sensor network applications require that the sensors in the network agree on the time. A global clock in a sensor system will help process and analyze the data correctly and predict future system behavior. For example, in the vehicle tracking application, each sensor may know the time when a vehicle is approaching. By matching the sensor location and sensing time, the sensor system may predict the vehicle moving direction and speed. Without a global agreement on time, the data from different sensors cannot be matched up. Other applications that need global clock synchronization include environment monitoring (for example, temperature), navigation guidance, and any other application that requires the coordination of locally sensed data and mobility. Clock synchronization may also help to conserve energy in a sensor network, by allowing a coordinated way to set nodes into sleeping mode. This leads to more complex communication since a node must compute when to wake up to receive a message.

In a paper just submitted to Infocom 2004 we discussed three methods for global synchronization in a sensor network: (1) the all-node-based method, (2) the cluster-based method, and (3) a fully localized diffusion-based method. The all-node-based method assumes the transmission time of a packet across a hop is the same for all nodes. It uses a packet to go around a cycle that is composed of all the nodes in the network and amortizes the packet transmission time on the cycle to each hop. This method does not scale well because it requires the nodes in the whole network to participate in the synchronization process at the same time. To address the scalability issue, we propose a hierarchical method. We use clusters to organize the whole network. The cluster head nodes are synchronized by using the first method and in each cluster the members are synchronized with the cluster head. These two methods are not localized; each synchronization process involves all the nodes in that network partition. To achieve full scalability, we propose a fully localized diffusion-based method with both synchronous and asynchronous implementations, in which each node exchanges and updates information locally with its neighbors. No global operations are required. In the synchronous rate-based algorithm, neighboring nodes exchange clock reading values proportional to their clock difference in a set order. To make our implementation more practical, we propose two asynchronous implementations, in which a node can synchronize with its neighbors at any time in any order. The asynchronous algorithms can also adapt to node failure, adverse communication channel, and node mobility.

Although our algorithms are aiming at solving the synchronization problem in a sensor network, they can be easily extended to the data aggregation problem, e.g., finding the average, highest, and lowest sensor data reading among all the sensors in the whole network. For example, in the all-node-based algorithm, the reading of a sensor can be attached to the message when the message used in the all-node-based synchronization is going along the cycle composed of all nodes. The sum of the readings (thus the average reading), the highest, and lowest reading over the whole network can be computed post hoc or on the fly. The diffusion-based algorithm is straightforward to be extended as well as we will see below.

25.2 Products

This project has produced many papers, in addition to working prototype software (running on Berkeley motes) for many of the algorithms.

Publications

[ABC+03] Javed Aslam, Zack Butler, Florin Constantin, Valentino Crespi, George Cybenko, and Daniela Rus. Tracking a moving object with a binary sensor network. In *Proceedings of the First International*

Conference on Embedded Networked Sensor Systems (SenSys), pages 150–161, Los Angeles, CA, November 2003. ACM Press.

[ALR03] Javed Aslam, Qun Li, and Daniela Rus. Three power-aware routing algorithms for sensor networks. *Wireless Communications and Mobile Computing*, 3(2):187–208, March 2003.

[BR03] Zack Butler and Daniela Rus. Event-based control for mobile sensors. *IEEE Pervasive Computing*, 2(4):34–42, Oct–Dec 2003.

[CPR03] Peter Corke, Ronald Peterson, and Daniela Rus. Networked robots: Flying robot navigation using a sensor net. In *Proceedings of the Eleventh International Symposium of Robotics Research (ISRR)*, Springer Tracts on Advanced Robotics (STAR). Springer-Verlag, October 2003.

[KSP+03] George Kantor, Sanjiv Singh, Ron Peterson, Daniela Rus, Aveek Das, Vijay Kumar, Guilherme Pereira, and John Spletzer. Distributed search and rescue with robot and sensor team. In *Proceedings of the Fourth International Conference on Field and Service Robotics*, pages 327–332. Sage Publications, July 2003.

[LAR01] Qun Li, Javed Aslam, and Daniela Rus. Online power-aware routing in wireless ad-hoc networks. In *Proceedings of the Seventh Annual International Conference on Mobile Computing and Networking*, pages 97–107. ACM Press, July 2001.

[LAR03] Qun Li, Javed Aslam, and Daniela Rus. Distributed energy-conserving routing protocols for sensor network. In *Proceedings of the 37th Hawaii International Conference on System Science*, January 2003.

[LDR03] Qun Li, Michael De Rosa, and Daniela Rus. Distributed algorithms for guiding navigation across a sensor network. In *Proceedings of the Ninth Annual International Conference on Mobile Computing and Networking*, pages 313–325, San Diego, September 2003. ACM Press.

[Li04] Qun Li. *Mobility and Communication in Sensor Networks*. PhD thesis, Dept. of Computer Science, Dartmouth College, August 2004.

[LPDR03] Qun Li, Ron Peterson, Michael DeRosa, and Daniela Rus. Reactive behavior in self-reconfiguring sensor network. *ACM Mobile Computing and Communications Review*, 7(1):56–68, January 2003.

[LPRR02] Qun Li, Ron Peterson, Michael De Rosa, and Daniela Rus. Reactive behavior in self-reconfiguring sensor networks. In *Proceedings of the Eighth Annual International Conference on Mobile Computing and Networking*. ACM Press, September 2002. Poster abstract; later published in MC2R.

[LR00] Qun Li and Daniela Rus. Sending messages to mobile users in disconnected ad-hoc wireless networks. In *Proceedings of the Sixth Annual International Conference on Mobile Computing and Networking (Mobicom)*, pages 44–55, Boston, August 2000. ACM Press.

[LR02] Qun Li and Daniela Rus. Message relay in disconnected ad-hoc networks. In *IEEE MASCOTS Workshop on Mobility and Wireless Access*, October 2002.

[LR03] Qun Li and Daniela Rus. Communication in disconnected ad-hoc networks using message relay. *Journal of Parallel and Distributed Computing*, 63(1):75–86, January 2003.

[LR04] Qun Li and Daniela Rus. Global clock synchronization in sensor networks. In *Proceedings of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, May 2004.

[LRR02] Qun Li, Michael De Rosa, and Daniela Rus. Distributed algorithms for guiding navigation across a sensor net. Technical Report TR2002-435, Dept. of Computer Science, Dartmouth College, October 2002.

[PR02] Ronald Peterson and Daniela Rus. Interacting with a sensor network. In *Proceedings of the Australian Conference for Robotics and Automation*, pages 105–110. Australian Robotics & Automation Association, November 2002.

25.3 Contributions

This work has demonstrated solutions to several important problems in the field: (1) a routing algorithm that supports mobility and disconnections and is provably correct; (2) a routing algorithm that optimizes power consumption; (3) a distributed mapping algorithm in perception space for sensor networks; and (5) a distributed application of sensor networks to navigation guidance and supporting algorithms. We have also contributed a new device for interacting with sensor networks and data from several physical experiments.

26 PKI and Trust (Sean Smith)

Our project focuses on issues of trust and computing: whose machine was carrying out whose calculation; what party was on the other end of the wire. Two themes dominate:

- **Trusted Computing Platforms.** In a distributed world, how can a stakeholder trust computation happening on a remote machine? Lacking protections against physical attacks (and, for that matter, even basic software attacks), the standard desktop will not suffice.
- **Public Key Infrastructure.** In a distributed world, how can a stakeholder verify the identity (or other relevant properties) of a remote party? Lacking shared secrets or even common organizational membership, standard "userid/password" methods will not suffice.

26.1 Activities and Findings

26.1.1 Trusted Computing

One part of this work lies in follow-ups to Prof. Smith's IBM 4758 work: a look at the underlying theory behind the outbound authentication design [Smi02, Smi04b], a 1000X improvement in DES speed [LS01], and a general retrospective [DLP+01].

Another part of this program has been to try *using* this platform to this worked. So far, students and Prof. Smith have used these secure coprocessor platforms to address problems such as:

- **Hardening Web Server Applications.** Why should you trust that remote Web server with your credit card number, when its operator might sell it on the black market? (The standard *secure sockets layer (SSL)* protocol for Web security only protects the channel, not what happens at the end.)

  M.S. student Shan Jiang prototyped applications of secure coprocessor technology for hardening Web servers against insider attack [JSM01, Smi01].

- **Armored Data Vaults.** Why should you trust law enforcement to abide by its own policy when it comes to accessing an archive of sensitive data?

  Senior Alex Iliev (who stayed on for a Ph.D.) prototyped applications for protecting private archive data from abuse [IS03b].

- **Practical Private Information Retrieval (PPIR).** How can a patron of a digital library trust that the library is not recording the name of each book he or she examines—or more subtle correlations and statistics?

  With an IBM colleague, Prof. Smith developed a scheme for using the 4758 platform for practical private information retrieval [SS01]—essentially, one cycles through the entire record set for each request. In his Ph.D. work under J. Freytag, Dmitri Asonov followed up with a two-step approach: an $O(N^2)$ step *shuffles* the N records; the kth request since the shuffle can then take $O(k)$ [AF03, Asn04].

  Alex Iliev and Prof. Smith then tried Dmitri's approach—an adaptation for previously theoretical *oblivious RAM* work [GO96]—for a Dartmouth-sized X.509 directory, but found the shuffle step would take several weeks. Instead, we used the dusty idea of Benes networks [Wak68] to perform an oblivious shuffle in $O(N \log N)$, bringing this preprocessing time down to a few hours [IS03a]. Subsequently, we've reduced the internal storage from $O(N \log N)$ bits to $O(k \log N)$ bits, and also added the ability for users to modify as well as read records [IS04].

In other work in this area, we've helped explore the potential of this hardware for securing online auctions [PSST02]; An in-press journal paper [IS05] summarizes our privacy-protecting applications.

TCPA/TCG. Platforms such as the IBM 4758 will never be ubiquitous. As a perhaps inevitable consequence of its high physical security, such a device is more expensive and far less powerful than a typical desktop. To

overcome these constraints, one might consider shrinking the physical security boundary to a one or two chips, embedding these chips in a commodity motherboard in such a way as to protect the broader platform, and making these features cheap enough so that the resulting platform might be ubiquitous (although sacrificing a degree of physical security along the way).

As an academic group, our lab is not in a position to cause such technology to penetrate the commercial market. However, an industrial consortium would be. The *Trusted Computing Platform Association (TCPA)*—now reformed as the *Trusted Computing Group (TCG)*—produced a series of specifications for a *Trusted Platform Module (TPM)*, a small chip to be added to a motherboard that participates in the boot process and provides a credential store keyed to a series of platform configuration registers. Apparently, the version 1.2 TPM is intended to participate with Intel's *LaGrande Technology (LT)* CPUs in supporting Microsoft's *Next Generation Secure Computing Base (NGSCB)*—except the 1.2 TPM is not publicly available, an implementation of the supporting *TCG Software Stack (TSS)* is not publicly available, LT is not available, and Microsoft has distanced itself from NGSCB. However, the 1.1b TPM has been shipping with commodity machines (such as IBM Netvistas and Thinkpads) for several years now, and its spec was public. What can we do?

First, we built the world's first (and only) open-source TCPA/TCG-based platform [Enf].

- In 2002, initially as his senior thesis, Rich MacDonald began digesting the spec and building a library to enable us to work with the TPM. In early summer 2003, as he was finishing testing, IBM Watson scooped us.

- However, with Ph.D. student John Marchesini and M.E.M. student Omen Wild, we used both libraries to design and prototype *Bear/Enforcer*, an architecture that uses the TPM to verify the integrity of a long-lived *core Linux security module (LSM)*, which in turn protects medium-lived file structure against a signed configuration file and protects short-lived data via an encrypted loopback filesystem. This work received a flurry of attention—more than 30K hits in the months after initial release, and over 1K sourcecode downloads since [MSWM03].

- Using Linux leaves our platform susceptible to attacks from malicious applications or a malicious root user. Senior Alex Barsamian and M.S. student Josh Stabiner addressed these problems by integrating Bear/Enforcer with the LSM provided by the NSA's *SE/Linux*.

Then, we used this platform for applications: [MSW+04]

- **Hardened Web Servers.** We revisited our 4758 project by installing Apache on the platform—using the TPM to bind the SSL private key to Apache, and to the site content, and to the OS that protects it. The indirection in Bear/Enforcer's integrity architecture lets us balance the mismatch between the long life of the server's keypair with the shorter lives of software components—the server does not need to get a new certificate each time Apache is upgraded.

- **Compartmented Attestation.** Using the SE/Linux compartments lets us balance the interests of a remote party, who wants to trust that an application X is running in a trustworthy way on a given machine, with the interests of the machine's owner, who doesn't want the party to know anything else happening on the machine.

- **OpenCA.** We used Bear/Enforcer to bind the private key of CA to its software—increasing the trustworthiness of the certificates it creates and the ability a parent or peer to cross-certify it.

- **OpenSSL.** In order to maximize impact of this work, we built an *OpenSSL engine* module that works with the TPM in our platform—easing the port of any application that uses OpenSSL for cryptography [Sta05].

Big Picture. The recent *IEEE Computer* column [Smi04a] summarizes my view on the use of hardware to enhance computer trust; the recent book [Smi05a] presents a broader view.

26.1.2 Public Key Infrastructure

Traditionally, the issues of computer security were implicitly framed in terms of a universal standard of goodness holding out against a world of adversaries. This notion persists today—in practices such as a user authenticating by providing a name and password (clearly, the service receiving these is trustworthy) and in terms such as "trusted computing."

The emerging information infrastructure requires a more complicated and nuanced model. Multiple parties spanning multiple organizations have various opinions about what they might trust, in what contexts. To be effective, the underlying mechanics for transmitting and expressing assertions about these parties needs to be able to provide the right parameters to accommodate this multiplicity of views.

Because it can enable secure communications between parties who do not share a secret *a priori*, public key cryptography is a natural and perhaps unique building block. For example, it can enable B to produce a digital signature on a message—and then enable A to verify later that B produced this signature. Achieving this vision in the real world, however, requires a *public key infrastructure (PKI)*: B must have a public/private keypair, A must know the binding the between B's public key and and some property of B, and both A and B must possess IT tools that enable effective use of these keys.

With colleagues from Computing Services and Computer Science, Prof. Smith helped found the *Dartmouth PKI Lab* to investigate these obstacles. As part of this work alone, Prof. Smith has served as principal investigator for over $1.7 million in grants (AT&T/Internet2, Mellon Foundation) enlisted the participation of colleagues from and the Department of Sociology at Dartmouth, served as the first program chair of the NIST/NIH/Internet2 *PKI Research Workshop* (we're up to #4 this year), and on the first program committee of *EuroPKI*, founded in response.

Under this leadership, the PKI Lab pursued both research into the missing pieces, as well as experimental deployment within the higher education community.

PKI Research.

Mental Models. One branch of this work explored whether, for basic applications the deployment community considered, the reality of what the systems were doing with the cryptography matched what the users and designers thought they were doing.

- **Server-side SSL.** Can a user A correctly distinguish what his or her computer is saying about the trustworthiness of a remote B?

  Web users rely on their browsers to display signals—such as the lock icon—to communicate the identity and security of the server at the other end. Two students and Prof. Smith extended Felten's classic work by discovering and demonstrated that the richness of what a browser will render on behalf of a remote server can permit a malicious server to send content that effectively mimics arbitrary aspects of the browser's user interface, *including these security signals* [YYS02].

- **Digital Signatures.** Can a user A conclude that, if their Office or email tools report a document was signed by B, that B was aware of and approved the virtual piece of paper that A sees? We've also been exploring the vulnerabilities of current implementations—and the fundamental limits of this technology. As part of the former, we've shown how commercial PKI packages can permit creation of documents that change in usefully malicious ways (such as an expense report whose numbers grow after the chair signs it) without invalidating the signature [KSA02]. (Jøsang [JDA02] concurrently published similar results—using an orthogonal set of techniques!)

- **Client-side SSL.** This variation of the standard SSL protocol uses identity PKI—based on a key pair at the browser—to let a server authenticate the user of a Web application. However, client-side SSL raises a new set of user interface and protocol issues: how does B know that her browser only uses

her private key for the service B wanted? Our work on *keyjacking* that suggests this problem is serious and subtle: in particular, the standard Windows/IE environment requires the entire machine to be the TCB [MSZ03, MSZ04].

This work led to speculation on the role of HCI for effective PKI, and an invitation to participate in the 2003 *ACM Workshop on Human-Computer Interaction and Security Systems*—the foundation of the new field of HCISEC [Smi03c].

We also initiated projects to address some of these problems.

- **Trusted Paths for Browsers.** Our Web spoofing work above demonstrated that popular browser interfaces cannot securely express the existence of an SSL channel or the identity of the server. We followed up that work by designing, prototyped, and tested an effective countermeasure within open-source Mozilla [YS02, YSA04].

- **Secure-Hardware-Enhanced MyProxy (SHEMP).** Our keyjacking work discussed above showed that standard desktops are not good places to keep private keys. Our Bear/Enforcer trusted computing work provides a way to increase assurance of ordinary desktops; the Grid community's *MyProxy* provides a foundation for authentication using temporary keypairs, and OASIS' *XACML* provides a standard way of expressing policies. In ongoing work, Ph.D.student John Marchesini is combining these tools to produce a way for users to employ proxy keypairs for signatures and encryption as well, limited by predefined policies geared toward the trustworthiness of the client platform [MS04b, MS04a].

Relying Parties. Another branch explored how we can ease the job of the relying party.

- **Virtual Hierarchies.** Within broader PKI trust architectures, relying parties must choose between the robustness of mesh architectures and the easy path construction of hierarchies. In early work (and as part of the Marianas project with David Nicol), Marchesini and Smith used the 4758 trusted computing platform and P2P to build *virtual hierarchies* for PKI that achieve the advantages of both approaches [MS02].

- **Distributed SEM.** In traditional PKI, a CA binds a public key to some property of the keyholder. Relying parties need to be able to determine whether or not the CA has *revoked* this binding. To simplify revocation checking in non-compromise scenarios, Boneh et al proposed the*Semi-Trusted Mediator (SEM)* approach [BDTW01]. Both a mediator and the user hold shares of the user's private key and must participate in the private key operation; the mediator can instantly revoke the key by deleting its each share.

  The initial SEM approach had problems with trust and scalability. As part of his senior thesis, Gabe Vanrenen and Smith designed and prototyped a way to distribute SEM by using a P2P network of mediators (to improve availability), hardware-enhanced trusted computing platforms (to improve trust), and threshold cryptography and strong forward secrecy (to mitigate damage from compromised mediators) [VS04]. We've since ported this onto our Bear/Enforcer trusted computing platform [VSM05].

Expressiveness. We've also explored the expressiveness of current PKI systems.

- **Greenpass.** In traditional PKI, a central, distinguished authority binds names to public keys. In many real-world scenarios, names may not be the appropriate parameter, and a central authority may not exist. In our Greenpass project, we explore this setting in practice—using lightweight SPKI/SDSI authorization grafted on to X.509 to permit local users to delegate access to guests, in a WLAN secured by EAP-TLS [GKS+04]. This projected attracted a $100K donation from Cisco, and also formed the initial inspiration of my Intel URC grant.

- In our paper at Allerton in 2004, we took a more thorough the mismatches between the authorization expressed by standard PKI tools and the authorization actually required by real-world scenarios [SMS04].

Performance. With Ph.D. student Meiyuan Zhao, we have also been exploring large-scale performance of PKI-based protocols, using parallel simulation.

- With Dave Nicol, we examined the performance impact of the signatures and verification that *S-BGP* uses to secure Internet routing path announcements. As part of this work, we discovered some ways of amortizing signatures that has less impact than the most optimistic S-BGP optimizations proposed, without their costs [NSZ04].
- Subsequently, we extended this analysis to examine origin authentication, certificate revocation, and recent aggregate signature proposals [ZSN05].
- In ongoing work sponsored by Sun Corporation, Meiyuan is using this same framework to examine the performance of certificate path discovery protocols designed by Sun's Internet Security Labs.

PKI Deployment Our lab, in conjunction with Dartmouth Computing Services, has also been carrying out practical deployment work.

- We've set up a college CA and made low-assurance certificates available to the campus population.
- We've modified the back-end Banner system to accept client-side PKI authentication, for campus information services such as registering for classes.
- In Fall 2005, we made higher-assurance certificates, housed on USB dongles, available to incoming freshmen; the college also required the Human Resources department to switch to these dongles.
- We're building and operating the *Higher Education Bridge CA (HEBCA)* chartered by EDUCAUSE [Hig]. HEBCA will facilitate inter-operation of individual academic PKIs through the US and Canada, and may end up the largest bridge CA in existence.

This deployment work is synergistic with research. For example, it contributed to the keyjacking and signature-hacking work; we are carrying out ongoing user studies of the dongle population; we hope to harness the the Bear/Enforcer OpenCA at remote universities to ease registration into HEBCA; we also hope to use HEBCA in inter-institutional non-identity PKI experiments.

26.2 Products

We have published one book [Smi05a] and one book chapter [CKST01]. We have published ten papers in refereed journals [DLP+01, IS05, MSZ04, NSZ04, PSST02, SS03, Smi01, Smi04b, SS01, YSA04]. We have published 19 papers in refereed conferences [AS04, GKS+04, IS03b, IS03a, IS04, JSM01, KSA02, LS01, MS02, MSZ03, MSW+04, NS04, SS04a, SS04b, Smi02, SMS04, VS04, YS02, ZSN05]. We have published seven other departments and columns [Smi05b, Smi04c, Smi04a, SS04c, Smi03a, Smi03c, Smi03b] and four additional technical reports [BSK04, MS05, MSWM03, YYS02].

Additionally, we have given more than four dozen invited talks.

This project has generated ten master's theses [Agr03, Ali03, Gof04, Jia01, Kai03, Kim04, Naz03, Sta05, Ye02] and eleven bachelor's theses [Bar02, Bar04, Ili01, Kan03, Kno01, Per03, Pow04, Ric01, Sel03, Van03, Wan04].

Publications

[Agr03] S. Agrawal. Investigation of Third Party Rights Service and Shibboleth Modification to Introduce the Service. Master's thesis, Dartmouth College Department of Computer Science, 2003.

[Ali03] Y. Ali. Adding Public Key Security to SSH. Master's thesis, Dartmouth College Department of Computer Science, 2003.

[AS04] Y. Ali and S.W. Smith. Flexible and Scalable Public Key Security for SSH. In *1st European PKI Workshop: Research and Applications*, pages 45–56. Springer-Verlag LNCS 3093, 2004. http://www.cs.dartmouth.edu/~sws/papers/ali.pdf.

[Bar02] M. Barreno. The Future of Cryptography under Quantum Computers. Senior thesis, Dartmouth College Department of Computer Science, 2002.

[Bar04] A. Barsamian. Software Compartmentalization and Attestation Using SELinux and the TPM/TCPA. Senior thesis, Dartmouth College Department of Computer Science, 2004.

[BSK04] K.-H. Baek, S.W. Smith, and D. Kotz. A Survey of WPA and 802.11i RSN Authentication Protocols. Technical Report TR2004-524, Dartmouth College, Computer Science, Hanover, NH, November 2004. ftp://ftp.cs.dartmouth.edu/TR/TR2004-524.pdf.

[CKST01] S. Chari, P. Kermani, S.W. Smith, and L. Tassiulas. Security Issues in M-Commerce: A Usage-Based Taxonomy. In *E-Commerce Agents: Marketplace Solutions, Security Issues, and Supply and Demand*, pages 264–283. Springer-Verlag LNCS 2033, April 2001. http://www.cs.dartmouth.edu/~sws/papers/ckst.pdf.

[GKS+04] N. Goffee, S.H. Kim, S.W. Smith, W. Taylor, M. Zhao, and J. Marchesini. Greenpass: Decentralized, PKI-based Authorization for Wireless LANs. In *3rd Annual PKI Research and Development Workshop Proceedings*, pages 26–41. NIST/NIH/Internet2; NISTIR 7122, April 2004. http://www.cs.dartmouth.edu/~sws/papers/greenpass-pki04-final.pdf.

[Gof04] N. Goffee. Greenpass Client Tools for Delegated Authorization in Wireless Networks. Master's thesis, Dartmouth College Department of Computer Science, 2004.

[Ili01] A. Iliev. An Armored Data Vault. Senior thesis, Dartmouth College Department of Computer Science, 2001.

[IS03a] A. Iliev and S.W. Smith. Privacy-Enhanced Credential Services. In *2nd Annual PKI Research Workshop Proceedings*. NIST/NIH/Internet2, April 2003. http://www.cs.dartmouth.edu/~sws/papers/ilsm03.pdf.

[IS03b] A. Iliev and S.W. Smith. Prototyping an Armored Data Vault: Rights Management for Big Brother's Computer. In *Privacy Enhancing Technologies—PET 2002*, pages 144–159. Springer-Verlag LNCS 2482, 2003. http://www.cs.dartmouth.edu/~sws/papers/pet02.pdf.

[IS04] A. Iliev and S.W. Smith. Private Information Storage with Logarithmic-space Secure Hardware. In *Information Security Management, Education, and Privacy*, pages 201–216. Kluwer, 2004. Proceedings of *i-NetSec 04: 3rd Working Conference on Privacy and Anonymity in Networked and Distributed Systems*. http://www.cs.dartmouth.edu/~sws/papers/is04.pdf.

[IS05] A. Iliev and S.W. Smith. Protecting User Privacy via Trusted Computing at the Server: Two Case Studies. *IEEE Security and Privacy*, 2005. In press.

[Jia01] S. Jiang. WebALPS Implementation and Performance Analysis: Using Trusted Co-servers to Enhance Privacy and Security of Web Interactions. Master's thesis, Dartmouth College Department of Computer Science, June 2001.

[JSM01] S. Jiang, S.W. Smith, and K. Minami. Securing Web Servers against Insider Attack. In *Seventeenth Annual Computer Security Applications Conference*, pages 265–276. IEEE Computer Society, 2001. http://www.cs.dartmouth.edu/~sws/papers/jsm.pdf.

[Kai03] K. Kain. Electronic Documents and Digital Signatures. Master's thesis, Dartmouth College Department of Computer Science, June 2003.

[Kan03] B.D. Kang. Strengthening Voice Authentication with Splicing Detection for User with Untrusted Clients. Senior thesis, Dartmouth College Department of Computer Science, 2003.

[Kim04] S.H. Kim. Greenpass RADIUS Tools for Delegated Authorization in Wireless Networks. Master's thesis, Dartmouth College Department of Computer Science, 2004.

[Kno01] E. Knop. Secure Public-Key Services for Web-Based Mail. Senior thesis, Dartmouth College Department of Computer Science, 2001.

[KSA02] K. Kain, S.W. Smith, and R. Asokan. Digital Signatures and Electronic Documents: A Cautionary Tale. In *Advanced Communications and Multimedia Security*, pages 293–307. Kluwer Academic Publishers, 2002. http://www.cs.dartmouth.edu/~sws/papers/cms02.pdf.

[LS01] M. Lindemann and S.W. Smith. Improving DES Coprocessor Throughput for Short Operations. In *Proceedings of the 10th USENIX Security Symposium*, pages 67–81, August 2001. http://www.cs.dartmouth.edu/~sws/papers/des.pdf.

[MS02] J. Marchesini and S.W. Smith. Virtual Hierarchies: An Architecture for Building and Maintaining Efficient and Resilient Trust Chains. In *Proceedings of the 7th Nordic Workshop on Secure IT Systems—NORDSEC 2002*. Karlstad University Studies, November 2002. http://www.cs.dartmouth.edu/~sws/papers/vh.pdf.

[MS05] J. Marchesini and S.W. Smith. SHEMP: Secure Hardware Enhanced MyProxy. Technical Report TR2005-532, Dartmouth College, Computer Science, Hanover, NH, February 2005. ftp://ftp.cs.dartmouth.edu/TR/TR2005-532.pdf.

[MSW+04] J. Marchesini, S.W. Smith, O. Wild, J. Stabiner, and A. Barsamian. Open-Source Applications of TCPA Hardware. In *20th Annual Computer Security Applications Conference*. IEEE Computer Society, December 2004. http://www.cs.dartmouth.edu/~sws/papers/acsac04.pdf.

[MSWM03] John Marchesini, Sean W. Smith, Omen Wild, and Rich MacDonald. Experimenting with TCPA/TCG Hardware, Or: How I Learned to Stop Worrying and Love The Bear. Technical Report TR2003-476, Dartmouth College, Computer Science, Hanover, NH, December 2003. This report supercedes TR2003-471 of August 2003. ftp://ftp.cs.dartmouth.edu/TR/TR2003-476.pdf.

[MSZ03] J. Marchesini, S.W. Smith, and M. Zhao. Keyjacking: Risks of the current client-side infrastructure. In *2nd Annual PKI Research Workshop Proceedings*. NIST/NIH/Internet2, April 2003. http://www.cs.dartmouth.edu/~sws/papers/keyjack.pdf.

[MSZ04] J. Marchesini, S.W. Smith, and M. Zhao. Keyjacking: the Surprising Insecurity of Client-side SSL. *Computers and Security*, 2004. In press. http://www.cs.dartmouth.edu/~sws/papers/kj04.pdf.

[Naz03] S. Nazareth. SPADE: SPKI/SDSI for Attribute Release Policies in a Distributed Environment. Master's thesis, Dartmouth College Department of Computer Science, 2003.

[NS04] S. Nazareth and S.W. Smith. Using SPKI/SDSI for Distributed Maintenance of Attribute Release Policies in Shibboleth. In *Proceedings of the IADIS International Conference WWW/Internet 2004*, volume 1, pages 218–226, October 2004. http://www.cs.dartmouth.edu/~sws/papers/nazareth04.pdf.

[NSZ04] D.M. Nicol, S. W. Smith, and M. Zhao. Evaluation of efficient security for BGP route announcements using parallel simulation. *Simulation Modelling Practice and Theory*, 12:187–216, 2004. http://dx.doi.org/10.1016/j.simpat.2003.10.003.

[Per03] M. Pereira. Trusted S/MIME Gateways. Senior thesis, Dartmouth College Department of Computer Science, 2003.

[Pow04] K. Powell. Testing the Greenpass Wireless Security System. Senior thesis, Dartmouth College Department of Computer Science, 2004.

[PSST02] A. Perrig, S.W. Smith, D. Song, and J.D. Tygar. SAM: A Flexible and Secure Auction Architecture using Trusted Hardware. *eJETA.org: The Electronic Journal for E-Commerce Tools and Applications*, 1, January 2002. http://www.cs.dartmouth.edu/~sws/papers/sam.pdf.

[Ric01] S. Richardson. Whom can you trust? Designing a System to Evaluate and Represent the Security State of Machines using PKI. Senior thesis, Dartmouth College Department of Computer Science, 2001.

[Sel03] P. Seligman. An Implementation of Machine Learning Algorithms to Detect Abnormal Data Access from Online Journal Archives. Senior thesis, Dartmouth College Department of Computer Science, 2003.

[Smi01] S.W. Smith. WebALPS: A Survey of E-Commerce Privacy and Security Applications. *ACM SIGecom Exchanges*, 2.3:27–34, September 2001. http://www.cs.dartmouth.edu/~sws/papers/acm.pdf.

[Smi02] S.W. Smith. Outbound Authentication for Programmable Secure Coprocessors. In *Computer Security—ESORICS 2002*, pages 72–89. Springer-Verlag LNCS 2502, October 2002. http://www.cs.dartmouth.edu/~sws/papers/esorics02.pdf.

[Smi03a] S.W. Smith. A Funny Thing Happened on the Way to the Marketplace. *IEEE Security and Privacy*, 1(6):74–78, November/December 2003. http://www.cs.dartmouth.edu/~sws/papers/marketplace.pdf.

[Smi03b] S.W. Smith. Fairy Dust, Secrets and the Real Worl. *IEEE Security and Privacy*, 1(1):89–93, January/February 2003. http://www.cs.dartmouth.edu/~sws/papers/fairydust.pdf.

[Smi03c] S.W. Smith. Humans in the Loop: Human-Computer Interaction and Security. *IEEE Security and Privacy*, 1(3):75–79, May/June 2003. http://www.cs.dartmouth.edu/~sws/papers/humans.pdf.

[Smi04a] S.W. Smith. Magic Boxes and Boots: Security in Hardware. *IEEE Computer*, 37(10):106–109, October 2004. http://www.cs.dartmouth.edu/~sws/papers/magic_boxes.pdf.

[Smi04b] S.W. Smith. Outbound Authentication for Programmable Secure Coprocessors. *International Journal of Information Security*, 3(1):28–41, 2004. http://www.springerlink.com/link.asp?id=tmej0glenmd2tfrm.

[Smi04c] S.W. Smith. Probing End-User IT Security Practices—via Homework. *The Educause Quarterly*, 27(4):68–71, November 2004. http://www.cs.dartmouth.edu/~sws/papers/eq.pdf.

[Smi05a] S.W. Smith. *Trusted Computing Platforms: Design and Applications*. Springer, 2005. http://www.springeronline.com/sgw/cda/frontpage/0,11855,4-148-22-36520721-0,00.html.

[Smi05b] S.W. Smith. Turing is from Mars, Shannon is from Venus: Computer Science and Computer Engineering, Separated by a Common Field. *IEEE Security and Privacy*, 2005. To appear.

[SMS04] S.W. Smith, C. Masone, and S. Sinclair. Expressing Trust in Distributed Systems: the Mismatch Between Tools and Reality. In *Forty-Second Annual Allerton Conference on Communication, Control, and Computing*, September 2004. Invited paper. http://www.cs.dartmouth.edu/~sws/papers/sms04.pdf.

[SS01] S.W. Smith and D. Safford. Practical Server Privacy Using Secure Coprocessors. *IBM Systems Journal*, 40:683–695, 2001. http://www.cs.dartmouth.edu/~sws/papers/sysj.pdf.

[SS03] A. Shubina and S. W. Smith. Using caching for browsing anonymity. *ACM SIGecom Exchanges*, 4.2:11–20, Summer 2003.

[SS04a] P. Seligman and S.W. Smith. Detecting Unauthorized Use in Online Journal Archives: A Case Study. In *Proceedings of the IADIS International Conference WWW/Internet 2004*, volume 1, pages 209–217, October 2004. http://www.cs.dartmouth.edu/~sws/papers/seligman04.pdf.

[SS04b] A. Shubina and S.W. Smith. Design and Prototype of a Coercion-Resistant, Verifiable Electronic Voting System. In *Proceedings of Second Annual Conference on Privacy, Security and Trust*, pages 29–39, October 2004. http://www.cs.dartmouth.edu/~sws/papers/ss04.pdf.

[SS04c] S.W. Smith and E. Spafford. Grand Challenges in Information Security: Process and Output. *IEEE Security and Privacy*, 2(1):69–71, January/February 2004. http://www.cs.dartmouth.edu/~sws/papers/challenges.pdf.

[Sta05] J. Stabiner. Providing Compartmented Hardware-Based Security on Commodity Machines. Master's thesis, Dartmouth College Department of Computer Science, 2005.

[Van03] G. Vanrenen. Distributed SEM: Extending SEM for Use on Distributed, Trusted Third Parties. Senior thesis, Dartmouth College Department of Computer Science, 2003.

[VS04] G. Vanrenen and S.W. Smith. Distributing Security-Mediated PKI. In *1st European PKI Workshop: Research and Applications*, pages 218–231. Springer-Verlag LNCS 3093, 2004. http://www.cs.dartmouth.edu/~sws/papers/gabe.pdf.

[VSM05] G. Vanrenen, S.W. Smith, and J. Marchesini. Distributing Security-Mediated PKI, January 2005. Revised and extended version, submitted for journal publication.

[Wan04] J. Wang. An Evaluation of Code Attestation in Trusted Computing Platforms. Senior thesis, Dartmouth College Department of Computer Science, 2004.

[Ye02] E. Ye. Building Trusted Paths for Web Browsers. Master's thesis, Dartmouth College Department of Computer Science, 2002.

[YS02] E. Ye and S.W. Smith. Trusted Paths for Browsers. In *Proceedings of the 11th USENIX Security Symposium*, August 2002. http://www.cs.dartmouth.edu/~sws/papers/usenix02.pdf.

[YSA04] E. Ye, S.W. Smith, and D. Anthony. Trusted Paths for Browsers. *ACM Transactions on Information System Security*, 2004. In press. http://www.cs.dartmouth.edu/~sws/papers/YSA.pdf.

[YYS02] E. Ye, Y. Yuan, and S.W. Smith. Web Spoofing Revisited: SSL and Beyond. Technical Report TR2002-417, Dartmouth College, Computer Science, Hanover, NH, February 2002. ftp://ftp.cs.dartmouth.edu/TR/TR2002-417.pdf.

[ZSN05] M. Zhao, S.W. Smith, and D. Nicol. Evaluating the Performance Impact of PKI on BGP Security. In *4th Annual PKI Research and Development Workshop*. NIST/NIH/Internet2, 2005. In press.

References

[AF03] D. Asonov and J. Freytag. Almost Optimal Private Information Retrieval. In *Privacy Enhancing Technologies—PET 2002*, pages 209–223. Springer-Verlag LNCS 2482, 2003.

[Asn04] D. Asnonov. *Querying Databases Privately: A New Approach to Private Information Retrieval*. Springer-Verlag LNCS 3128, 2004.

[BDTW01] D. Boneh, X. Ding, G. Tsudik, and C.M. Wong. A Method for Fast Revocation of Public Key Certificates and Security Capabilities. In *Proceedings of the 10th USENIX Security Symposium*, pages 297–308, 2001.

[DLP+01] J. Dyer, M. Lindemann, R. Perez, R. Sailer, S.W. Smith, L.van Doorn, and S. Weingart. Building the IBM 4758 Secure Coprocessor. *IEEE Computer*, 34:57–66, October 2001. http://www.cs.dartmouth.edu/~sws/papers/comp01.pdf.

[DPSL99] J. Dyer, R. Perez, S.W. Smith, and M. Lindemann. Application Support Architecture for a High-Performance, Programmable Secure Coprocessor. In *22nd National Information Systems Security Conference*, October 1999. http://www.cs.dartmouth.edu/~sws/papers/nimp.pdf.

[Enf] Enforcer Homepage. http://enforcer.sourceforge.net/.

[FJS91] M. Furst, J. Jackson, and S. Smith. Improved Learning of AC0 Functions. In *4th ACM Annual Workshop on Computational Learning Theory*, 1991. http://www.cs.dartmouth.edu/~sws/papers/colt91.pdf.

[GO96] O. Goldreich and R. Ostrovsky. Software Protection and Simulation on Oblivious RAMs. *Journal of the ACM*, 43(3):431–473, 1996.

[GSTY96] H. Gobioff, S.W. Smith, J.D. Tygar, and B.S. Yee. Smart Cards in Hostile Environments. In *Proceedings of the 2nd USENIX Workshop on Electronic Commerce*, pages 23–28, 1996. http://www.cs.dartmouth.edu/~sws/papers/ec96card.pdf.

[Hig] Higher Education Bridge Certification Authority. http://www.educause.edu/hebca/.

[JDA02] A. Jøsang, D.Povey, and A.Ho. What You See is Not Always What You Sign. In *Proceedings of AUUG2002*, September 2002.

[MS04a] J. Marchesini and S.W. Smith. Modeling Public Key Infrastructure in the Real World, November 2004. Submitted for publication.

[MS04b] J. Marchesini and S.W. Smith. Secure Hardware Enhanced MyProxy. Technical Report TR2004-525, Dartmouth College, Computer Science, Hanover, NH, November 2004. ftp://ftp.cs.dartmouth.edu/TR/TR2004-525.pdf.

[SA98] S.W. Smith and V. Austel. Trusting Trusted Hardware: Towards a Formal Model for Programmable Secure Coprocessors. In *Proceedings of the 3rd USENIX Workshop on Electronic Commerce*, August 1998. http://www.cs.dartmouth.edu/~sws/papers/ec98.pdf.

[SJ96] S.W. Smith and D.B. Johnson. Minimizing Timestamp Size for Completely Asynchronous Optimistic Recovery with Minimal Rollback. In *15th IEEE Symposium on Reliable Distributed Systems*, October 1996. http://www.cs.dartmouth.edu/~sws/papers/srds96.pdf.

[SJT95] S.W. Smith, D.B. Johnson, and J.D. Tygar. Completely Asynchronous Optimistic Recovery with Minimal Rollbacks. In *25th IEEE International Symposium on Fault-Tolerant Computing*, 1995. http://www.cs.dartmouth.edu/~sws/papers/ftcs95.pdf.

[Smi96] S.W. Smith. Secure Coprocessing Applications and Research Issues. Technical Report Los Alamos Unclassified Release LA-UR-96-2805, Los Alamos National Laboratory, August 1996. http://www.cs.dartmouth.edu/~sws/papers/lanl.pdf.

[SP96] S.W. Smith and P. Pedersen. Organizing Electronic Services into Security Taxonomies. In *Proceedings of the 2nd USENIX Workshop on Electronic Commerce*, 1996. http://www.cs.dartmouth.edu/~sws/papers/ec96tax.pdf.

[SPW98] S.W. Smith, E. Palmer, and S. Weingart. Using a High-Performance, Programmable Secure Coprocessor. In *Financial Cryptography, Second International Conference, FC'98*, pages 73–89. Springer-Verlag LNCS 1465, 1998. http://www.cs.dartmouth.edu/~sws/papers/fc98.pdf.

[SPWA99] S.W. Smith, R. Perez, S.H. Weingart, and V. Austel. Validating a High-Performance, Programmable Secure Coprocessor. In *22nd National Information Systems Security Conference*, October 1999. http://www.cs.dartmouth.edu/~sws/papers/nfips.pdf.

[ST91] S.W. Smith and J.D. Tygar. Signed Vector Timestamps: A Secure Protocol for Partial Order Time. Technical Report CMU-CS-93-116, CMU, Computer Science, Pittsburgh, PA, 1991. Version of February 1993. http://www.cs.dartmouth.edu/~sws/papers/signed.pdf.

[ST94] S.W. Smith and J.D. Tygar. Security and Privacy for Partial Order Time. In *ISCA Seventh International Conference on Parallel and Distributed Computing Systems*, October 1994. http://www.cs.dartmouth.edu/~sws/papers/pdcs94.pdf.

[Sta05] J. Stabiner. Providing Compartmented Hardware-based Security on Commodity Machines. Master's thesis, Dartmouth College, Department of Computer Science, 2005. To appear.

[SW99] S.W. Smith and S. Weingart. Building a High-Performance, Programmable Secure Coprocessor. *Computer Networks*, 31:831–860, April 1999. http://www.cs.dartmouth.edu/~sws/papers/sw.pdf.

[Wak68] A. Waksman. A Permutation Network. *Journal of the ACM*, 15(1):159–163, 1968.

Software Some online materials:

1. Early demonstrations of Web SSL spoofing, and countermeasure patches to Mozilla: http://www.cs.dartmouth.edu/~pkilab/demos
2. World's first (and, to date, only) open-source TCPA/TCG-based platform: http://enforcer.sourceforge.net
3. Numerous training and deployment resources for academic PKI: http://www.dartmouth.edu/~pkilab/

26.3 Contributions

We've enabled trusted computing research by creating open source platforms that can (and have been) used for other projects.

With the book and our PKILab outreach materials, we've created informational resources that reach beyond the lab and into general academia. With HEBCA, we're creating infrastructure resources as well.

27 Algorithms (Cliff Stein)

Professor Stein's research is in the design and analysis of algorithms, combinatorial optimization, network algorithms, scheduling and computational biology.

Professor Stein moved to Columbia University in 2001, so this report covers only his time at Dartmouth College.

27.1 Activities and Findings

Much of Stein's work was on scheduling algorithms to minimize average completion time, average flow time, and on optimizing these metrics while simultaneously minimizing schedule length. Phillips, Stein and Wein [PSW98] introduced the idea of using a relaxed schedule to get an ordering in which to schedule jobs. This paper started a flurry of activity on average completion time problems. Inspired by this paper, Hall et. al [HSSW97], in some very elegant work, showed how to use various linear programming formulations, together with modified ordering rules, to get small-constant factor approximation algorithms for scheduling problems with precedence constraints and release dates and unrelated machines. Following this work, Stein worked on improved approximation factors for many average completion time problems [CPS+96]. In addition, in joint work with Chekuri et. al.[CMNS01], Stein developed the idea of introducing randomization into the ordering rule. This yields a $e/(e-1) \approx 1.58$-approximation algorithm for the "one machine, jobs arriving over time" case. (This idea was independently introduced by Goemans[Goe97] for a related problem.) The ideas mentioned above have recently been used by several researchers, in various combinations, and with various additional ideas, to achieve even better bounds for several average completion time problems [Sch95, MSS96, CH99, CS97, Goe97, SS97b, SS97a, CM99, MQW97, GQS+99]. In addition, recent experimental work has shown that the ideas introduced yield improved performance in practice[SRW98].

Stein then showed further theoretical improvements. In an 11 authored FOCS 99 paper[ABC+99], he presented the first known polynomial time approximation schemes for several average completion time problems. Results include PTASs for the case of identical parallel machines and a constant number of unrelated machines with and without preemption allowed. The PTASs are also efficient. For most variants, the running time for a $(1+\epsilon)$–approximation on an instance with n jobs and m machines is $O(n \log n)$ for each fixed ϵ, and for all variants the running time's dependence on n is a fixed polynomial independent of ϵ and m.

For the case of a single machine and release dates, with a student, Clint Hepner, he [HS01] have showed that the approximation scheme actually shows promise of performing well in practice. This is particularly interesting because the running time of a PTAS is usually so large that implementing the algorithm is not feasible. For this reason, it is rare that a PTAS yields an efficient implementation of an algorithm. The scheduling PTAS discussed in this paper has a worst-case running time of $O(n \lg n + (1/\epsilon^5)!)$. Although the constant term is large, the fact that it is not coupled with the $n \lg n$ term makes it plausible that the algorithm can be implemented to run quickly. They have compared the new PTAS to several other exact, heuristic, and approximation algorithms, and compare the performance of the PTAS to these other algorithms. They also find better algorithms based on the ideas developed in the PTAS. Given the limited practical success of implementing approximation schemes, lessons in the implementation that may be applicable to other approximation schemes. These results show that although the PTAS when implemented strictly, does not perform very well, it is possible to modify the implementation to develop strong heuristic algorithms that perform well in practice. These method for doing so focuses on the impact of rounding and enumeration and thus may be applicable to approximation schemes for other problems as well.

Stein has also studied schedules which are simultaneously near-optimal for two criteria. This follows on work done jointly with Wein [SW97], and appears in a SODA 99 paper [ARSY99], the undergraduate thesis of April Rasala [Ras99], a paper [RTSU02]. In these works, they give improved bounds on existence theorems, in a very general setting, for bicriterion scheduling problems, where the objective are average completion time and schedule length. The techniques involve mapping the schedule to a probability density function and using analytical techniques to solve a min-max problem, where the function to be evaluated involves integrating a

density function. They then used these techniques to prove existence theorems for a large class of objectives including those based on completion time, flow time, lateness and the number of on-time jobs.

In addition to this new work, Stein, together with David Karger and Joel Wein, wrote the chapter on scheduling for the CRC Handbook on Algorithms[KSW98].

Stein has also worked on graph algorithms.

An IPCO 2001 paper with a student David Wagner [SW01], introduced a variant of the travelling salesman problem in which the metric is to minimize the number of turns in the tour, given that the input points are in the Euclidean plane. It is motivated by applications in robotics and in the movement of other heavy machinery: for many such devices turning is an expensive operation. For the general case of an arbitrary set of n points in the Euclidean plane, they give a logarithmic approximation algorithm. For the case when the lines of the tour are restricted to being either horizontal or vertical, they give a 2-approximation algorithm. With the further restriction that no two points are allowed to have the same x- or y-coordinate, they give an algorithm that finds a tour which makes at most two turns more than the optimal tour. Thus the approximation algorithm has an additive, rather than a multiplicative error bound. Beyond the additive error bound, the algorithm for this problem introduces several interesting algorithmic techniques for decomposing sets of points in the Euclidean plane.

Together with a student, Stavros Kolliopoulos, Stein worked the unsplittable flow problem and related disjoint path problems[KS97, KS98, KS99].

In the single-source unsplittable flow problem we are given a graph G, a source vertex s and a set of sinks $t_1, \ldots, t_k$ with associated demands. We seek a single s-t_i flow path for each i so that the demands are satisfied and the total flow routed across any edge e is bounded by its capacity c_e. The problem is an NP-hard variant of max flow and a generalization of single-source disjoint paths with applications to scheduling, load balancing and virtual-circuit routing problems. In a significant development, Kleinberg gave recently constant-factor approximation algorithms for several natural optimization versions of the problem [Kle96]. In this work Stein and Kolliopoulos give a generic framework that yields simpler algorithms and significant improvements upon the constant factors. This framework, with appropriate subroutines, applies to all optimization versions previously considered and treats in a unified manner directed and undirected graphs.

To give a flavor of the results, consider minimizing maximum congestion, i.e. the maximum ratio over all edges e of the flow through e over the capacity c_e. This metric was a primary testbed for randomized rounding techniques and has been studied extensively. They give a simple $(4 + o(1))$-approximation algorithm for both directed and undirected graphs. The previously known bounds were 16 for the directed and 8.25 for the undirected case. This approach also gives the first constant factor approximation for minimum cost unsplittable flow on directed graphs and improves considerably upon the approximation ratio for the minimum cost version on undirected graphs.

They also extended this work to some more general problems in approximating disjoint-path type problems. In the *edge(vertex)-disjoint path* problem we are given a graph G and a set $\mathcal{T}$ of connection requests. Every connection request in $\mathcal{T}$ is a vertex pair (s_i, t_i), $1 \leq i \leq K$. The objective is to connect a maximum number of the pairs via edge(vertex)-disjoint paths. The edge-disjoint path problem can be generalized to the *multiple-source unsplittable flow* problem where commodity i has a demand ρ_i and every edge e a capacity u_e. All these problems are NP-hard and have a multitude of applications in areas such as routing, scheduling and bin packing.

Although the edge- and vertex-disjoint path problems, and more recently the unsplittable flow generalization, have been extensively studied, they remain notoriously hard to approximate with a bounded performance guarantee. For example, even for the simple edge-disjoint paths problem, no $\omega(\sqrt{|E|})$-approximation algorithm is known. Moreover some of the best existing approximation ratios are obtained through sophisticated and non-standard randomized rounding schemes. They give new algorithms that lead to the first approximation algorithm for the general unsplittable flow problem. For this problem, even with weights on the commodities, they obtain an approximation ratio that matches, to within logarithmic factors, the $O(\sqrt{|E|})$ approximation ratio for the simple edge-disjoint paths problem. In addition to this result and to improved bounds for several other disjoint-path problems, the techniques simplify and unify the derivation of many existing approximation results.

They use two basic techniques. First, they propose simple greedy algorithms for edge- and vertex-disjoint paths and second, they propose the use of a framework based on *packing integer programs* for more general problems such as unsplittable flow. A packing integer program is of the form maximize $c^T \cdot x$, subject to $Ax \leq b$, $A, b, c \geq 0$. As part of our tools they develop improved approximation algorithms for a class of packing integer programs.

Stein and Kolliopoulos have also performed experimental work on the single-source unsplittable flow problem[KS99]. They have introduced various heuristics and shown how to distinguish between them experimentally. They have also addressed the problem of generating instances that are hard for the unsplittable flow problem.

Stein also worked on computational biology. That work is described in the section by Bruce Donald.

27.2 Products

Publications

[ABC+99] F. Afrati, E. Bampis, C. Chekuri, D. Karger, C. Kenyon, S. Khanna, I. Milis, M. Queyranne, M. Skutella, C. Stein, and M. Sviridenko. Approximation schemes for minimizing average weighted completion time with release dates. In *Proceedings of the 40th Annual Symposium on Foundations of Computer Science*, pages 32–43, 1999.

[ARSY99] J. A. Aslam, A. Rasala, C. Stein, and N. Young. Improved bicriteria existence theorems for scheduling. In *Proceedings of the 10th ACM-SIAM Symposium on Discrete Algorithms*, pages 846–847, 1999.

[CH99] F. Chudak and D. Hochbaum. A half-integral linear programming relaxation for scheduling precedence constrained jobs on a single machine. *Operations Research Letters*, 25:199–204, 1999.

[CM99] C. Chekuri and R. Motwani. Precedence constrained scheduling to minimize weighted completion time on a single machine. *Discrete Applied Mathematics*, 98:29–39, 1999.

[CMNS01] C. Chekuri, R. Motwani, B. Natarajan, and C. Stein. Approximation techniques for average completion time scheduling. *SIAM Journal on Computing*, 31(1):146–166, 2001.

[CPS+96] S. Chakrabarti, C. A. Phillips, A. S. Schulz, D. B. Shmoys, C. Stein, and J. Wein. Improved scheduling algorithms for minsum criteria. In F. Meyer auf der Heide and B. Monien, editors, *Automata, Languages and Programming*, number 1099 in Lecture Notes in Computer Science. Springer, Berlin, 1996. Proceedings of the 23rd International Colloquium (ICALP'96).

[CS97] F. A. Chudak and D. B. Shmoys. Approximation algorithms for precedence-constrained scheduling problems on parallel machines that run at different speeds. In *Proceedings of the 8th ACM-SIAM Symposium on Discrete Algorithms*, 1997.

[Goe97] M. Goemans. Improved approximation algorithms for scheduling with release dates. In *Proceedings of the 8th ACM-SIAM Symposium on Discrete Algorithms*, pages 591–598, 1997.

[GQS+99] M. Goemans, M. Queyranne, A. Schulz, M. Skutella, and Y. Wang. Single machine scheduling with release dates. Preprint, 1999.

[HS01] C. Hepner and C. Stein. Implementation of a PTAS for scheduling with release dates. In *Proceedings of ALENEX*, 2001.

[HSSW97] L. A. Hall, A. S. Schulz, D. B. Shmoys, and J. Wein. Scheduling to minimize average completion time: Off-line and on-line approximation algorithms. *Mathematics of Operations Research*, 22:513–544, August 1997.

[Kle96] J. M. Kleinberg. Single-source unsplittable flow. In *Proceedings of the 37th Annual Symposium on Foundations of Computer Science*, pages 68–77, October 1996.

[KS97] S. G. Kolliopoulos and C. Stein. Improved approximation algorithms for unsplittable flow problems. In *Proceedings of the 38th Annual Symposium on Foundations of Computer Science*, pages 426–436, 1997.

[KS98] S. G. Kolliopoulos and C. Stein. Approximating disjoint-path problems using greedy algorithms and packing integer programs. In *Proceedings of the 6th Conference on Integer Programming and Combinatorial Optimization*, pages 153–168, 1998.

[KS99] S. G. Kolliopoulos and C. Stein. Experimental evaluation of approximation algorithms for single source unsplittable flow. In *Proceedings of the 7th Conference on Integer Programming and Combinatorial Optimization*, 1999.

[KSW98] David Karger, Cliff Stein, and Joel Wein. *CRC Handbook on Algorithms*, chapter Scheduling Algorithms. CRC Press, 1998.

[MQW97] F. Margot, M. Queyranne, and Y. Wang. Decompositions, network flows and a precendece constrained single machine scheduling problem. talk by M. Queyranne at IMPS 97, 1997.

[MSS96] R. H. Möhring, M. W. Schäffter, and A. S. Schulz. Scheduling jobs with communication delays: Using infeasible solutions for approximation. In J. Diaz and M. Serna, editors, *Algorithms – ESA'96*, volume 1136 of *Lecture Notes in Computer Science*, pages 76 – 90. Springer, Berlin, 1996. Proceedings of the 4th Annual European Symposium on Algorithms.

[PSW98] C. Phillips, C. Stein, and J. Wein. Minimizing average completion time in the presence of release dates. *Mathematical Programming*, 82:199–223, 1998.

[Ras99] April Rasala. Existence theorems for scheduling to meet two objectives. Undergraduate Thesis, 1999. Dartmouth College, Dept. of Computer Science.

[RTSU02] A. Rasala, E. Torng, C. Stein, and P. Uthaisombut. Existence theorems, lower bound and algorithms for scheduling to meet two objectives. In *Proceedings of the 13th ACM-SIAM Symposium on Discrete Algorithms*, 2002.

[Sch95] A.S. Schulz. Scheduling to minimize total weighted completion time: performance guarantees of lp based heuristics and lower bounds. Technical Report 474/1995, Technical University of Berlin, 1995.

[SRW98] Martin W.P. Savelsbergh, R.N.Uma, and Joel Wein. An experimental study of LP-based scheduling heuristics. In *Proceedings of the 9th ACM-SIAM Symposium on Discrete Algorithms*, pages 453–461, 1998.

[SS97a] A. S. Schulz and M. Skutella. Random–based scheduling: New approximations and LP lower bounds. In J. Rolim, editor, *Randomization and Approximation Techniques in Computer Science*, volume 1269 of *LNCS*, pages 119 – 133. Springer, Berlin, 1997. Proceedings of the International Workshop RANDOM'97.

[SS97b] A. S. Schulz and M. Skutella. Scheduling–LPs bear probabilities: Randomized approximations for min–sum criteria. In R. Burkard and G. Woeginger, editors, *Algorithms – ESA'97*, volume 1284 of *LNCS*, pages 416 – 429. Springer, Berlin, 1997. Proceedings of the 5th Annual European Symposium on Algorithms.

[SW97] C. Stein and J. Wein. On the existence of schedules that are near-optimal for both makespan and total weighted completion time. *Operations Research Letters*, 21:115–122, 1997.

[SW01] C. Stein and D. Wagner. Approximation algorithms for the minimum bends traveling salesman problem. In *Proceedings of the 8th Conference on Integer Programming and Combinatorial Optimization*, pages 406–421, 2001.

In addition to these papers, Professor Stein released software package to find minimum cuts in graphs.[17]

Professor Stein was a co-author of the second edition of *Introduction to Algorithms*,[18] the leading textbook in the field.

27.3 Contributions

The Contributions are described in the activites and findings section.

Human resources. April Rasala, an undergraduate major in computer science, was a Senior Honors Thesis student with Professor Stein. During that time, she won the CRA outstanding undergraduate award. She just got her Ph.D. from MIT in March 2005.

[17] http://www.cs.dartmouth.edu/~cliff/code.html
[18] http://mitpress.mit.edu/algorithms

28 Technical Report series

The department disseminates many of its research findings through its Technical Report series, available at http://www.cs.dartmouth.edu/reports.

In this section we list all of the technical reports produced by the department. Although some of these TRs may also be listed in the preceding topical sections, we include them as another view into the productivity of our department. All of these are produced, in whole or in part, by the faculty and students of this department. Some of them represent our Ph.D, M.S., and undergraduate theses.

Publications

[ABK+04] Javed Aslam, Sergey Bratus, David Kotz, Ron Peterson, Daniela Rus, and Brett Tofel. The Kerf toolkit for intrusion analysis. Technical Report TR2004-493, Dartmouth College, Computer Science, Hanover, NH, March 2004.

[Abr00] Karolyn A. Abram. Registration of Images with Dissimilar Contrast using a Hybrid Method Employing Correlation and Mutual Information. Technical Report TR2000-369, Dartmouth College, Computer Science, Hanover, NH, June 2000.

[Agr03] Sanket Agrawal. Investigation of Third Party Rights Service and Shibboleth Modification to Introduce the Service . Technical Report TR2003-463, Dartmouth College, Computer Science, Hanover, NH, June 2003.

[AM00] Javed A. Aslam and Mark Montague. Bayes Optimal Metasearch: A Probabilistic Model for Combining the Results of Multiple Retrieval Systems. Technical Report TR2000-382, Dartmouth College, Computer Science, Hanover, NH, December 2000.

[Art00] John C. Artz. Personal Radio. Technical Report TR2000-372, Dartmouth College, Computer Science, Hanover, NH, June 2000.

[Bap99] Lauren M. Baptist. Two Algorithms for Performing Multidimensional, Multiprocessor, Out-of-Core FFTs. Technical Report PCS-TR99-350, Dartmouth College, Computer Science, Hanover, NH, June 1999.

[Bar02] Marco A. Barreno. The Future of Cryptography Under Quantum Computers. Technical Report TR2002-425, Dartmouth College, Computer Science, Hanover, NH, July 2002.

[BKKSD00] Chris Bailey-Kellogg, John J. Kelley, Clifford Stein, and Bruce Randall Donald. Reducing Mass Degeneracy in SAR by MS by Stable Isotopic Labeling. Technical Report TR2000-362, Dartmouth College, Computer Science, Hanover, NH, February 2000.

[BKR99] Jonathan Bredin, David Kotz, and Daniela Rus. Mobile-Agent Planning in a Market-Oriented Environment. Technical Report PCS-TR99-345, Dartmouth College, Computer Science, Hanover, NH, May 1999.

[BKR01] Chris Bailey-Kellogg and Naren Ramakrishnan. Ambiguity-Directed Sampling for Qualitative Analysis of Sparse Data from Spatially-Distributed Physical Systems. Technical Report TR2001-384, Dartmouth College, Computer Science, Hanover, NH, January 2001.

[BKWK+99] Chris Bailey-Kellogg, Alik Widge, John J. Kelley, Marcelo J. Berardi, John H. Bushweller, and Bruce Randall Donald. The NOESY Jigsaw: Automated Protein Secondary Structure and Main-Chain Assignment from Sparse, Unassigned NMR Data. Technical Report PCS-TR99-358, Dartmouth College, Computer Science, Hanover, NH, October 1999.

[BMI+99] Jonathan Bredin, Rajiv T. Maheswaran, Cagri Imer, Tamer Basar, David Kotz, and Daniela Rus. A Game-Theoretic Formulation of Multi-Agent Resource Allocation. Technical Report PCS-TR99-360, Dartmouth College, Computer Science, Hanover, NH, October 1999.

[BR03] Zack Butler and Daniela Rus. Distributed planning and control for modular robots with unit-compressible modules. Technical Report TR2003-462, Dartmouth College, Computer Science, Hanover, NH, June 2003.

[Bre01] Jonathan L. Bredin. Market-based Control of Mobile-agent Systems. Technical Report TR2001-408, Dartmouth College, Computer Science, Hanover, NH, June 2001.

[BSK04] Kwang-Hyun Baek, Sean W. Smith, and David Kotz. A Survey of WPA and 802.11i RSN Authentication Protocols. Technical Report TR2004-524, Dartmouth College, Computer Science, Hanover, NH, November 2004.

[CC03] Geeta Chaudhry and Thomas H. Cormen. Stupid Columnsort Tricks. Technical Report TR2003-444, Dartmouth College, Computer Science, Hanover, NH, April 2003.

[CCSR02] Valentino Crespi, George Cybenko, Massimo Santini, and Daniela Rus. Decentralized Control for Coordinated flow of Multi-Agent Systems. Technical Report TR2002-414, Dartmouth College, Computer Science, Hanover, NH, January 2002.

[CG00] Ezra E. K. Cooper and Robert S. Gray. An Economic CPU-Time Market for D'Agents. Technical Report TR2000-375, Dartmouth College, Computer Science, Hanover, NH, June 2000.

[Cha99] James D. Chalfant. Parallel DaSSF Discrete-Event Simulation without Shared Memory. Technical Report PCS-TR99-346, Dartmouth College, Computer Science, Hanover, NH, June 1999.

[Cha04] Geeta Chaudhry. Parallel Out-of-Core Sorting: The Third Way. Technical Report TR2004-517, Dartmouth College, Computer Science, Hanover, NH, September 2004.

[CHC03] Geeta Chaudhry, Elizabeth A. Hamon, and Thomas H. Cormen. Relaxing the Problem-Size Bound for Out-of-Core Columnsort. Technical Report TR2003-445, Dartmouth College, Computer Science, Hanover, NH, April 2003.

[Che04] Guanling Chen. Solar: Building A Context Fusion Network for Pervasive Computing. Technical Report TR2004-514, Dartmouth College, Computer Science, Hanover, NH, August 2004.

[Chy00] Debbie O. Chyi. An Infrastructure for a Mobile-Agent System that Provides Personalized Services to Mobile Devices. Technical Report TR2000-370, Dartmouth College, Computer Science, Hanover, NH, May 2000.

[CK00] Guanling Chen and David Kotz. A Survey of Context-Aware Mobile Computing Research. Technical Report TR2000-381, Dartmouth College, Computer Science, Hanover, NH, November 2000.

[CK01] Guanling Chen and David Kotz. Supporting Adaptive Ubiquitous Applications with the SOLAR System. Technical Report TR2001-397, Dartmouth College, Computer Science, Hanover, NH, May 2001.

[CK02a] Guanling Chen and David Kotz. Context Aggregation and Dissemination in Ubiquitous Computing Systems. Technical Report TR2002-420, Dartmouth College, Computer Science, Hanover, NH, February 2002.

[CK02b] Guanling Chen and David Kotz. Solar: A pervasive-computing infrastructure for context-aware mobile applications. Technical Report TR2002-421, Dartmouth College, Computer Science, Hanover, NH, February 2002.

[CK04a] Guanling Chen and David Kotz. A Case Study of Four Location Traces. Technical Report TR2004-490, Dartmouth College, Computer Science, Hanover, NH, February 2004.

[CK04b] Guanling Chen and David Kotz. Application-controlled loss-tolerant data dissemination. Technical Report TR2004-488, Dartmouth College, Computer Science, Hanover, NH, February 2004.

[CK04c] Guanling Chen and David Kotz. Dependency management in distributed settings. Technical Report TR2004-495, Dartmouth College, Computer Science, Hanover, NH, March 2004.

[Cre02] Valentino Crespi. Exact formulae for the Lovasz Theta Function of sparse Circulant Graphs. Technical Report TR2002-438, Dartmouth College, Computer Science, Hanover, NH, November 2002.

[DBKKS99] Bruce Randall Donald, Chris Bailey-Kellogg, John J. Kelley, and Clifford Stein. SAR by MS for Functional Genomics (Structure-Activity Relation by Mass Spectrometry). Technical Report PCS-TR99-359, Dartmouth College, Computer Science, Hanover, NH, October 1999.

[DH99] Bruce Randall Donald and Frederick Henle. Using Haptic Vector Fields for Animation Motion Control. Technical Report PCS-TR99-353, Dartmouth College, Computer Science, Hanover, NH, May 1999.

[Dub04] Nikita E Dubrovsky. Mobile Agents Simulation with DaSSF. Technical Report TR2004-499, Dartmouth College, Computer Science, Hanover, NH, June 2004.

[Ess02] Kobby Essien. Analysis of Protein Sequences Using Time Frequency and Kolmogorov-Smirnov Methods. Technical Report TR2002-431, Dartmouth College, Computer Science, Hanover, NH, May 2002.

[Far00] Hany Farid. Reconstructing Ancient Egyptian Tombs. Technical Report TR2000-383, Dartmouth College, Computer Science, Hanover, NH, December 2000.

[Far01] Hany Farid. Detecting Steganographic Messages in Digital Images. Technical Report TR2001-412, Dartmouth College, Computer Science, Hanover, NH, September 2001.

[Far04a] Hany Farid. Creating and Detecting Doctored and Virtual Images: Implications to The Child Pornography Prevention Act . Technical Report TR2004-518, Dartmouth College, Computer Science, Hanover, NH, September 2004.

[Far04b] Hany Farid. Discrete-Time Fractional Differentiation from Integer Derivatives . Technical Report TR2004-528, Dartmouth College, Computer Science, Hanover, NH, December 2004.

[Fin01] Jeremy T. Fineman. Optimizing the Dimensional Method for Performing Multidimensional, Multiprocessor, Out-of-Core FFTs. Technical Report TR2001-402, Dartmouth College, Computer Science, Hanover, NH, June 2001.

[Fit02] Robert C. Fitch. Heterogeneous Self-Reconfiguring Robotics. Technical Report TR2002-436, Dartmouth College, Computer Science, Hanover, NH, November 2002.

[Fit04] Robert C. Fitch. Heterogeneous Self-Reconfiguring Robotics. Technical Report TR2004-519, Dartmouth College, Computer Science, Hanover, NH, September, 2004.

[GGK+01] Arne Grimstrup, Robert S. Gray, David Kotz, Thomas Cowin, Greg Hill, Niranjan Suri, Daria Chacon, and Martin Hofmann. Write Once, Move Anywhere: Toward Dynamic Interoperability of Mobile Agent Systems. Technical Report TR2001-411, Dartmouth College, Computer Science, Hanover, NH, July 2001.

[GKCR00] Robert S. Gray, David Kotz, George Cybenko, and Daniela Rus. Mobile Agents: Motivations and State-of-the-Art Systems. Technical Report TR2000-365, Dartmouth College, Computer Science, Hanover, NH, April 2000.

[GKN+04] Robert S. Gray, David Kotz, Calvin Newport, Nikita Dubrovsky, Aaron Fiske, Jason Liu, Christopher Masone, Susan McGrath, and Yougu Yuan. Outdoor Experimental Comparison of Four Ad Hoc Routing Algorithms. Technical Report TR2004-511, Dartmouth College, Computer Science, Hanover, NH, June 2004.

[GKP+01] Robert S. Gray, David Kotz, Ronald A. Peterson, Peter Gerken, Martin Hofmann, Daria Chacon, Greg Hill, and Niranjan Suri. Mobile-Agent versus Client/Server Performance: Scalability in an Information-Retrieval Task. Technical Report TR2001-386, Dartmouth College, Computer Science, Hanover, NH, January 2001.

[Gof04] Nicholas C. Goffee. Greenpass Client Tools for Delegated Authorization in Wireless Networks. Technical Report TR2004-509, Dartmouth College, Computer Science, Hanover, NH, June 2004.

[Ham03] Elizabeth A. Hamon. Enhancing Asynchronous Parallel Computing. Technical Report TR2003-460, Dartmouth College, Computer Science, Hanover, NH, June 2003.

[HH02] Heng Huang and Chris Hawblitzel. Proofs of Soundness and Strong Normalization for Linear Memory Types. Technical Report TR2002-437, Dartmouth College, Computer Science, Hanover, NH, November 2002.

[HHW04] Chris Hawblitzel, Heng Huang, and Lea Wittie. Composing a Well-Typed Region. Technical Report TR2004-521, Dartmouth College, Computer Science, Hanover, NH, October 2004.

[HK99] Jon Howell and David Kotz. An Access-Control Calculus for Spanning Administrative Domains. Technical Report PCS-TR99-361, Dartmouth College, Computer Science, Hanover, NH, November 1999.

[HK00a] Jon Howell and Keith Kotay. Landmarks for absolute localization. Technical Report TR2000-364, Dartmouth College, Computer Science, Hanover, NH, March 2000.

[HK00b] Jon Howell and David Kotz. A Formal Semantics for SPKI. Technical Report TR2000-363, Dartmouth College, Computer Science, Hanover, NH, March 2000.

[HK03] Tristan Henderson and David Kotz. Problems with the Dartmouth wireless SNMP data collection . Technical Report TR2003-480, Dartmouth College, Computer Science, Hanover, NH, December 2003.

[HKA04] Tristan Henderson, David Kotz, and Ilya Abyzov. The Changing Usage of a Mature Campus-wide Wireless Network. Technical Report TR2004-496, Dartmouth College, Computer Science, Hanover, NH, March 2004.

[How00a] Jon Howell. Naming and sharing resources across administrative boundaries (errata). Technical Report TR2000-380, Dartmouth College, Computer Science, Hanover, NH, May 2000.

[How00b] Jon Howell. Naming and sharing resources across administrative boundaries (Volume I). Technical Report TR2000-378, Dartmouth College, Computer Science, Hanover, NH, May 2000.

[How00c] Jon Howell. Naming and sharing resources across administrative boundaries (Volume II). Technical Report TR2000-379, Dartmouth College, Computer Science, Hanover, NH, May 2000.

[HRKM02] Dennis M. Healy, Daniel N. Rockmore, Peter J. Kostelec, and Sean S. B. Moore. FFTs for the 2-Sphere - Improvements and Variations. Technical Report TR2002-419, Dartmouth College, Computer Science, Hanover, NH, March 2002.

[HWH03] Heng Huang, Lea Wittie, and Chris Hawblitzel. Formal Properties of Linear Memory Types. Technical Report TR2003-468, Dartmouth College, Computer Science, Hanover, NH, August 2003.

[IA99] Jeffrey D. Isaacs and Javed A. Aslam. Investigating Measures for Pairwise Document Similarity. Technical Report PCS-TR99-357, Dartmouth College, Computer Science, Hanover, NH, June 1999.

[IS03] Alex Iliev and Sean Smith. Privacy-enhanced credential services. Technical Report TR2003-442, Dartmouth College, Computer Science, Hanover, NH, February 2003.

[Iyi01] Mehmet Iyigun. DaSSFNet: An Extension to DaSSF for High-Performance Network Modeling. Technical Report TR2001-405, Dartmouth College, Computer Science, Hanover, NH, June 2001.

[Jia01] Shan Jiang. WebALPS Implementation and Performance Analysis: Using Trusted Co-servers to Enhance Privacy and Security of Web Interactions. Technical Report TR2001-399, Dartmouth College, Computer Science, Hanover, NH, June 2001.

[JP03] Prasad Jayanti and Srdjan Petrovic. Efficient and Practical Constructions of LL/SC Variables. Technical Report TR2003-446, Dartmouth College, Computer Science, Hanover, NH, June 2003.

[JP04] Prasad Jayanti and Srdjan Petrovic. Efficient Wait-Free Implementation of Multiword LL/SC Variables. Technical Report TR2004-523, Dartmouth College, Computer Science, Hanover, NH, October 2004.

[JSM01] Shan Jiang, Sean Smith, and Kazuhiro Minami. Securing Web Servers against Insider Attack. Technical Report TR2001-410, Dartmouth College, Computer Science, Hanover, NH, July 2001.

[KCG+00] David Kotz, George Cybenko, Robert S. Gray, Guofei Jiang, Ronald A. Peterson, Martin O. Hofmann, Daria A. Chacon, Kenneth R. Whitebread, and James Hendler. Performance Analysis of Mobile Agents for Filtering Data Streams on Wireless Networks. Technical Report TR2000-377, Dartmouth College, Computer Science, Hanover, NH, October 2000.

[KE02a] David Kotz and Kobby Essien. Analysis of a Campus-wide Wireless Network. Technical Report TR2002-432, Dartmouth College, Computer Science, Hanover, NH, September 2002.

[KE02b] David Kotz and Kobby Essien. Characterizing Usage of a Campus-wide Wireless Network. Technical Report TR2002-423, Dartmouth College, Computer Science, Hanover, NH, March 2002.

[KGR02] David Kotz, Robert S. Gray, and Daniela Rus. Future Directions for Mobile-Agent Research. Technical Report TR2002-415, Dartmouth College, Computer Science, Hanover, NH, January 2002.

[Kha01a] Ammar Khalid. A Directory Infrastructure to Support Mobile Services. Technical Report TR2001-391, Dartmouth College, Computer Science, Hanover, NH, June 2001.

[Kha01b] Michael G. Khankin. TCP/IP Implementation within the Dartmouth Scalable Simulation Framework. Technical Report TR2001-407, Dartmouth College, Computer Science, Hanover, NH, June 2001.

[Kid01] Eric Kidd. Efficient Compression of Generic Function Dispatch Tables. Technical Report TR2001-404, Dartmouth College, Computer Science, Hanover, NH, June 2001.

[Kim04] Sung Hoon Kim. Greenpass RADIUS Tools for Delegated Authorization in Wireless Networks. Technical Report TR2004-510, Dartmouth College, Computer Science, Hanover, NH, June 2004.

[KJG+00] David Kotz, Guofei Jiang, Robert S. Gray, George Cybenko, and Ronald A. Peterson. Performance Analysis of Mobile Agents for Filtering Data Streams on Wireless Networks. Technical Report TR2000-366, Dartmouth College, Computer Science, Hanover, NH, May 2000.

[KNE03] David Kotz, Calvin Newport, and Chip Elliott. The mistaken axioms of wireless-network research. Technical Report TR2003-467, Dartmouth College, Computer Science, Hanover, NH, July 2003.

[KNG+04] David Kotz, Calvin Newport, Robert S. Gray, Jason Liu, Yougu Yuan, and Chip Elliott. Experimental evaluation of wireless simulation assumptions. Technical Report TR2004-507, Dartmouth College, Computer Science, Hanover, NH, June 2004.

[Kol99] Marisa E. Kolodny. Computers, Art and Smart Rooms: A Smart Picture Frame that Senses the Weather and Genetically Evolves Images. Technical Report PCS-TR99-354, Dartmouth College, Computer Science, Hanover, NH, June 1999.

[Kot04] David Kotz. Technological Implications for Privacy. Technical Report TR2004-505, Dartmouth College, Computer Science, Hanover, NH, June 2004.

[Kru04] Eric G. Krupski. PPL: a Packet Processing Language. Technical Report TR2004-508, Dartmouth College, Computer Science, Hanover, NH, June 2004.

[Kun99] David H. Kung. Improved Computer Detection and Mapping of Cerebral Oxygenation. Technical Report PCS-TR99-349, Dartmouth College, Computer Science, Hanover, NH, June, 1999.

[Lah02] Sebastien M. Lahaie. Information-theoretic Bounds on the Training and Testing Error of Boosting. Technical Report TR2002-428, Dartmouth College, Computer Science, Hanover, NH, May 2002.

[Lat01] David D. Latham. An Empirical Study of Training and Testing Error in Boosting. Technical Report TR2001-394, Dartmouth College, Computer Science, Hanover, NH, June 2001.

[LBKAD03] Ryan H. Lilien, Chris Bailey-Kellogg, Amy A. Anderson, and Bruce R. Donald. A Subgroup Algorithm to Identify Cross-Rotation Peaks Consistent with Non-Crystallographic Symmetry. Technical Report TR2003-481, Dartmouth College, Computer Science, Hanover, NH, December 2003.

[LD03a] Chris J. Langmead and Bruce R. Donald. 3D-Structural Homology Detection via Unassigned Residual Dipolar Couplings. Technical Report TR2003-439, Dartmouth College, Computer Science, Hanover, NH, January 2003.

[LD03b] Christopher J. Langmead and Bruce R. Donald. An Improved Nuclear Vector Replacement Algorithm for Nuclear Magnetic Resonance Assignment. Technical Report TR2004-494, Dartmouth College, Computer Science, Hanover, NH, September 2003.

[LD03c] Christopher J. Langmead and Bruce R. Donald. High-Throughput 3D Homology Detection via NMR Resonance Assignment. Technical Report TR2004-487, Dartmouth College, Computer Science, Hanover, NH, September 2003.

[Lee03] Clara E. Lee. Persistence and Prevalence in the Mobility of Dartmouth Wireless Network Users. Technical Report TR2003-455, Dartmouth College, Computer Science, Hanover, NH, May 2003.

[Len03] Chris Lentz. 802.11b Wireless Network Visualization and Radiowave Propagation Modeling . Technical Report TR2003-451, Dartmouth College, Computer Science, Hanover, NH, June 2003.

[LFD02] Ryan H. Lilien, Hany Farid, and Bruce R. Donald. Probabilistic Disease Classification of Expression-Dependent Proteomic Data from Mass Spectrometry of Human Serum. Technical Report TR2002-434, Dartmouth College, Computer Science, Hanover, NH, October 2002.

[Lif99] Artem Lifschitz. An Environment for the Facilitation of Robotic Programming. Technical Report PCS-TR99-356, Dartmouth College, Computer Science, Hanover, NH, June 1999.

[LNT01] Xiaowen Liu, David M. Nicol, and King Tan. Lock-free Scheduling of Logical Processes in Parallel Simulation. Technical Report TR2001-385, Dartmouth College, Computer Science, Hanover, NH, January 2001.

[LRR02] Qun Li, Michael De Rosa, and Daniela Rus. Distributed Algorithms for Guiding Navigation across a Sensor Network. Technical Report TR2002-435, Dartmouth College, Computer Science, Hanover, NH, October 2002.

[LSD04] Ryan H. Lilien, Mohini Sridharan, and Bruce R Donald. Identification of Novel Small Molecule Inhibitors of Core-Binding Factor Dimerization by Computational Screening against NMR Molecular Ensembles. Technical Report TR2004-492, Dartmouth College, Computer Science, Hanover, NH, March 2004.

[Lyu04] Siwei Lyu. Mercer Kernels for Object Recognition with Local Features. Technical Report TR2004-520, Dartmouth College, Computer Science, Hanover, NH, October 2004.

[Mar00] David B. Martin. Depth from Flash. Technical Report TR2000-373, Dartmouth College, Computer Science, Hanover, NH, June 2000.

[Mar01] Aidan S. Marcuss. EcomRISK.org : A site to classify and organize the risks of performing business on the Internet. Technical Report TR2001-403, Dartmouth College, Computer Science, Hanover, NH, June 2001.

[Mas02] Christopher P. Masone. Role Definition Language (RDL): A Language to Describe Context-Aware Roles. Technical Report TR2002-426, Dartmouth College, Computer Science, Hanover, NH, May 2002.

[Mat01] Arun Mathias. SmartReminder: A Case Study on Context-Sensitive Applications. Technical Report TR2001-392, Dartmouth College, Computer Science, Hanover, NH, June 2001.

[Maz04] Paul J. Mazzuca. Access Control in a Distributed Decentralized Network: An XML Approach to Network Security using XACML and SAML. Technical Report TR2004-506, Dartmouth College, Computer Science, Hanover, NH, Spring 2004.

[MK02] Kazuhiro Minami and David Kotz. Controlling access to pervasive information in the "Solar" system. Technical Report TR2002-422, Dartmouth College, Computer Science, Hanover, NH, February 2002.

[MK04] Kazuhiro Minami and David Kotz. Secure Context-sensitive Authorization. Technical Report TR2004-529, Dartmouth College, Computer Science, Hanover, NH, December 2004.

[Mon02] Mark H. Montague. Metasearch: Data Fusion for Document Retrieval. Technical Report TR2002-424, Dartmouth College, Computer Science, Hanover, NH, May 2002.

[MS02] John C. Marchesini and Sean W. Smith. Virtual Hierarchies - An Architecture for Building and Maintaining Efficient and Resilient Trust Chains. Technical Report TR2002-416, Dartmouth College, Computer Science, Hanover, NH, February 2002.

[MS04] John Marchesini and Sean W. Smith. Secure Hardware Enhanced MyProxy. Technical Report TR2004-525, Dartmouth College, Computer Science, Hanover, NH, November 2004.

[MS05] John Marchesini and Sean Smith. SHEMP: Secure Hardware Enhanced MyProxy. Technical Report TR2005-532, Dartmouth College, Computer Science, Hanover, NH, February 2005.

[MSMW03] Rich MacDonald, Sean W. Smith, John Marchesini, and Omen Wild. Bear: An Open-Source Virtual Secure Coprocessor based on TCPA. Technical Report TR2003-471, Dartmouth College, Computer Science, Hanover, NH, August 2003.

[MSWM03] John Marchesini, Sean W. Smith, Omen Wild, and Rich MacDonald. Experimenting with TCPA/TCG Hardware, Or: How I Learned to Stop Worrying and Love The Bear. Technical Report TR2003-476, Dartmouth College, Computer Science, Hanover, NH, December 2003.

[MSZ13] John Marchesini, Sean W. Smith, and Meiyuan Zhao. Keyjacking: The Surprising Insecurity of Client-side SSL. Technical Report TR2004-489, Dartmouth College, Computer Science, Hanover, NH, February 13,.

[MSZ03] John C. Marchesini, Sean W. Smith, and Meiyuan Zhao. Keyjacking: Risks of the Current Client-side Infrastructure. Technical Report TR2003-443, Dartmouth College, Computer Science, Hanover, NH, February 2003.

[MT01] G. Ayorkor Mills-Tettey. Mobile Voice Over IP (MVOIP): An Application-level Protocol. Technical Report TR2001-390, Dartmouth College, Computer Science, Hanover, NH, June 2001.

[Mun00] Martin Mundhenk. The complexity of planning with partially-observable Markov decision processes. Technical Report TR2000-376, Dartmouth College, Computer Science, Hanover, NH, June 2000.

[Naz03] Sidharth P. Nazareth. SPADE: SPKI/SDSI for Attribute Release Policies in a Distributed Environment . Technical Report TR2003-453, Dartmouth College, Computer Science, Hanover, NH, May 2003.

[New04] Calvin Newport. Simulating mobile ad hoc networks: a quantitative evaluation of common MANET simulation models. Technical Report TR2004-504, Dartmouth College, Computer Science, Hanover, NH, June 2004.

[NS04] Sidharth Nazareth and Sean Smith. Using SPKI/SDSI for Distributed Maintenance of Attribute Release Policies in Shibboleth. Technical Report TR2004-485, Dartmouth College, Computer Science, Hanover, NH, January 2004.

[NSZ03] David M. Nicol, Sean W. Smith, and Meiyuan Zhao. Efficient Security for BGP Route Announcements. Technical Report TR2003-440, Dartmouth College, Computer Science, Hanover, NH, May 2003.

[NY03] David M. Nicol and Guanhau Yan. Discrete-Event Fluid Modeling of Background TCP Traffic. Technical Report TR2003-454, Dartmouth College, Computer Science, Hanover, NH, June 2003.

[OK02] Ron Oldfield and David Kotz. Using the Emulab network testbed to evaluate the Armada I/O framework for computational grids. Technical Report TR2002-433, Dartmouth College, Computer Science, Hanover, NH, September 2002.

[Old03] Ron A. Oldfield. Efficient I/O for Computational Grid Applications. Technical Report TR2003-459, Dartmouth College, Computer Science, Hanover, NH, May 2003.

[Pea99] Matthew D. Pearson. Fast Out-of-Core Sorting on Parallel Disk Systems. Technical Report PCS-TR99-351, Dartmouth College, Computer Science, Hanover, NH, June 1999.

[Pec04a] Joseph E. Pechter. Enhancing Expressiveness of Speech through Animated Avatars for Instant Messaging and Mobile Phones. Technical Report TR2004-503, Dartmouth College, Computer Science, Hanover, NH, June 2004.

[Pec04b] William Pechter. Synchronizing Keyframe Facial Animation to Multiple Text-to-Speech Engines and Natural Voice with Fast Response Time. Technical Report TR2004-501, Dartmouth College, Computer Science, Hanover, NH, June 2004.

[Per03] Mindy J. Pereira. Trusted S/MIME Gateways. Technical Report TR2003-461, Dartmouth College, Computer Science, Hanover, NH, May 2003.

[PF01] Senthil Periaswamy and Hany Farid. Differential Elastic Image Registration. Technical Report TR2001-413, Dartmouth College, Computer Science, Hanover, NH, September 2001.

[PF04] Alin C. Popescu and Hany Farid. Exposing Digital Forgeries by Detecting Duplicated Image Regions . Technical Report TR2004-515, Dartmouth College, Computer Science, Hanover, NH, August 2004.

[Pop04] Alin C. Popescu. Statistical Tools for Digital Image Forensics. Technical Report TR2005-531, Dartmouth College, Computer Science, Hanover, NH, December 2004.

[Pow04] Kimberly S. Powell. Testing the Greenpass Wireless Security System. Technical Report TR2004-512, Dartmouth College, Computer Science, Hanover, NH, June 2004.

[Pre03] Brian J. Premore. An Analysis of Convergence Properties of the Border Gateway Protocol Using Discrete Event Simulation. Technical Report TR2003-452, Dartmouth College, Computer Science, Hanover, NH, May 2003.

[Ras99] April M. Rasala. Existence Theorems for Scheduling to Meet Two Objectives. Technical Report PCS-TR99-347, Dartmouth College, Computer Science, Hanover, NH, June 1999.

[Ric03] Evan W. Richardson. An Evaluation of the Impact of Models for Radio Propagation on the Simulation of 802.11b Wireless Networks. Technical Report TR2003-450, Dartmouth College, Computer Science, Hanover, NH, June 2003.

[Rin01] Michael F. Ringenburg. Applying the Vector Radix Method to Multidimensional, Multiprocessor, Out-of-Core Fast Fourier Transforms. Technical Report TR2001-388, Dartmouth College, Computer Science, Hanover, NH, March 2001.

[Rin04] Rachel B. Ringel. Efficient Wait-Free Implementation of Atomic Multi-Word Buffer. Technical Report TR2004-498, Dartmouth College, Computer Science, Hanover, NH, June, 2004.

[Rob01] Jeremy I. Robin. Fastab: Solving the Pitch to Notation Problem. Technical Report TR2001-406, Dartmouth College, Computer Science, Hanover, NH, May 2001.

[Ros03] Michael De Rosa. Power Conservation in the Network Stack of Wireless Sensors. Technical Report TR2003-458, Dartmouth College, Computer Science, Hanover, NH, June 2003.

[SFMS03] Li Shen, James Ford, Fillia Makedon, and Andrew Saykin. A Surface-based Approach for Classification of 3D Neuroanatomic Structures. Technical Report TR2003-464, Dartmouth College, Computer Science, Hanover, NH, June 2003.

[SGK+04] Sean Smith, Nicholas C. Goffee, Sung Hoon Kim, Punch Taylor, Meiyuan Zhao, and John Marchesini. Greenpass: Flexible and Scalable Authorization for Wireless Networks. Technical Report TR2004-484, Dartmouth College, Computer Science, Hanover, NH, January 2004.

[Sim03] Daniel F. Simola. Discovery, Visualization and Analysis of Gene Regulatory Sequence Elements in Genomes. Technical Report TR2003-456, Dartmouth College, Computer Science, Hanover, NH, May 2003.

[SKJH04] Libo Song, David Kotz, Ravi Jain, and Xiaoning He. Evaluating next-cell predictors with extensive Wi-Fi mobility data. Technical Report TR2004-491, Dartmouth College, Computer Science, Hanover, NH, February 2004.

[Smi01] Sean W. Smith. Outbound Authentication for Programmable Secure Coprocessors. Technical Report TR2001-401, Dartmouth College, Computer Science, Hanover, NH, March 2001.

[SS03] Anna M. Shubina and Sean W. Smith. Using caching for browsing anonymity. Technical Report TR2003-470, Dartmouth College, Computer Science, Hanover, NH, July 2003.

[Ste01a] Thomas B. Stephens. Improving a Brokering System for Linking Distributed Simulations. Technical Report TR2001-389, Dartmouth College, Computer Science, Hanover, NH, June 2001.

[Ste01b] Pablo Stern. Measuring early usage of Dartmouth's wireless network. Technical Report TR2001-393, Dartmouth College, Computer Science, Hanover, NH, June 2001.

[Sto04] Geoffrey H. Stowe. A Secure Network Node Approach to the Policy Decision Point in Distributed Access Control. Technical Report TR2004-502, Dartmouth College, Computer Science, Hanover, NH, June 2004.

[SW00] Cliff Stein and David P. Wagner. Approximation Algorithms for the Minimum Bends Traveling Salesman Problem. Technical Report TR2000-367, Dartmouth College, Computer Science, Hanover, NH, April 2000.

[Tan03] King Y. Tan. On the Complexity of Implementing Certain Classes of Shared Objects. Technical Report TR2003-475, Dartmouth College, Computer Science, Hanover, NH, November 2003.

[Tor03] Lisa A. Torrey. An Active Learning Approach to Efficiently Ranking Retrieval Engines. Technical Report TR2003-449, Dartmouth College, Computer Science, Hanover, NH, May 2003.

[VK04] Darren Erik Vengroff and David Kotz. A Holesome File System. Technical Report TR2004-497, Dartmouth College, Computer Science, Hanover, NH, May 2004.

[Von99] Marsette A. Vona. A Two Dimensional Crystalline Atomic Unit Modular Self-reconfigurable Robot. Technical Report PCS-TR99-348, Dartmouth College, Computer Science, Hanover, NH, June 1999.

[Wan04] Jue Wang. Performance Evaluation of a Resource Discovery Service. Technical Report TR2004-513, Dartmouth College, Computer Science, Hanover, NH, October, 2004.

[WCK04] Jue Wang, Guanling Chen, and David Kotz. A meeting detector and its applications. Technical Report TR2004-486, Dartmouth College, Computer Science, Hanover, NH, March 2004.

[Wei03] Edward Wei. Using Low Level Linear Memory Management for Type-Preserving Mark-Sweep Garbage Collector. Technical Report TR2003-465, Dartmouth College, Computer Science, Hanover, NH, June 2003.

[Wha99] Jason M. Whaley. An Application of Word Sense Disambiguation to Information Retrieval. Technical Report PCS-TR99-352, Dartmouth College, Computer Science, Hanover, NH, June 1999.

[Whi02a] A. Abram White. Performance and Interoperability In Solar. Technical Report TR2002-427, Dartmouth College, Computer Science, Hanover, NH, June 2002.

[Whi02b] A. Abram White. XSLT and XQuery as Operator Languages. Technical Report TR2002-429, Dartmouth College, Computer Science, Hanover, NH, June 2002.

[Wit04] Lea Wittie. Type-Safe Operating System Abstractions. Technical Report TR2004-526, Dartmouth College, Computer Science, Hanover, NH, June 2004.

[Won01a] Tiffany M. Wong. An Implementation of Object-Oriented Program Transformation for Thought-Guided Debugging . Technical Report TR2001-395, Dartmouth College, Computer Science, Hanover, NH, June 2001.

[Won01b] Tiffany M. Wong. Implementing a Database Information System for an Electronic Baseball Scorecard . Technical Report TR2001-396, Dartmouth College, Computer Science, Hanover, NH, June 2001.

[Ye02] Eileen Zishuang Ye. Building Trusted Paths for Web Browsers. Technical Report TR2002-430, Dartmouth College, Computer Science, Hanover, NH, May 2002.

[YLD03] Anthony K. Yan, Christopher J. Langmead, and Bruce Randall Donald. A Probability-Based Similarity Measure for Saupe Alignment Tensors with Applications to Residual Dipolar Couplings in NMR Structural Biology. Technical Report TR2003-474, Dartmouth College, Computer Science, Hanover, NH, October 2003.

[You99] Neal E. Young. Greedy Approximation Algorithms for K-Medians by Randomized Rounding. Technical Report PCS-TR99-344, Dartmouth College, Computer Science, Hanover, NH, March 1999.

[YS02] Eileen Ye and Sean Smith. Trusted Paths for Browsers: An Open-Source Solution to Web Spoofing. Technical Report TR2002-418, Dartmouth College, Computer Science, Hanover, NH, February 2002.

[YYS01] Yougu Yuan, Eileen Zishuang Ye, and Sean W. Smith. Web Spoofing 2001. Technical Report TR2001-409, Dartmouth College, Computer Science, Hanover, NH, July 2001.

[YYS02] Eileen Ye, Yougu Yuan, and Sean Smith. Web Spoofing Revisited: SSL and Beyond. Technical Report TR2002-417, Dartmouth College, Computer Science, Hanover, NH, February 2002.